Data Analysis Portfolio

Prepared by: - Akash S

Table of Contents

Professional Background	3
What is Data Analytics	4
Data Analysis Process	5
Instagram User Analytics	8
Operation and Metric Analytics	15
Hiring Process Analysis	31
IMDb Movie Analysis	35
Bank Loan Case Study	42
Impact of Car Features on Price and Profitability	62
ABC Call Volume Trend	73

Professional Background

I am currently pursuing a Master's degree in Data Analytics, building on a solid foundation with a Bachelor's degree in Computer Applications (BCA), where I secured a CGPA of 7.76. My academic journey and practical experiences have equipped me with a diverse skill set essential for the field of data analytics.

Skills:

- Data Analysis: Proficient in analysing complex datasets to extract actionable insights.

- Python: Experienced in using Python for data manipulation, analysis, and visualization.

- SQL: Skilled in querying databases to retrieve and analyse data efficiently.

- Microsoft Excel: Advanced capabilities in data handling, analysis, and visualization.

- Tableau: Competent in creating interactive and insightful data visualizations.

- Microsoft PowerBI: Experienced in developing comprehensive and dynamic business intelligence reports.

In addition to my technical skills, I have also worked as a professional photographer and videographer. This experience has honed my attention to detail, creativity, and ability to capture and present visual information effectively.

Combining my analytical expertise with my creative skills in photography and videography allows me to approach problems from unique perspectives and develop innovative solutions. This blend of talents makes me well-equipped to tackle diverse challenges in data analytics and beyond.

What is Data Analysis

Data analysis is a critical process that involves inspecting, cleaning, transforming, and modelling data to discover meaningful information and support decision-making. It is a multidisciplinary field that combines various statistical and computational techniques to extract insights from data. The importance of data analysis cannot be overstated. With the massive amount of data being generated by individuals, businesses, and governments on a daily basis, the need for skilled data analysts has never been greater. Organizations that can effectively collect, store, analyse, and interpret data can gain valuable insights into customer behaviour, market trends, and operational efficiency, giving them a competitive edge in their respective industries.

The data analysis process typically involves several steps. The first step is data collection, where data is gathered from various sources, such as databases, surveys, or sensors. Once the data is collected, it is cleaned and pre-processed to remove any errors or inconsistencies. The data is then organized and transformed to prepare it for analysis. The next step is data exploration, where data analysts use various statistical techniques to identify patterns, trends, and correlations in the data. This may involve creating visualizations, such as charts or graphs, to help identify key insights.

In conclusion, data analysis is a critical process that plays a vital role in modern organizations. By leveraging data analysis techniques, organizations can gain valuable insights into customer behaviour, market trends, and operational efficiency, giving them a competitive edge in their respective industries. Data analysts play a crucial role in this process, using their specialized skills and expertise to transform raw data into meaningful insights that drive business decisions.

Module - 1

Data Analytics Process

Description:

We use Data Analytics in everyday life without even knowing it.
For eg : Going to a market to buy something.

· **Plan**

We first decide which things I need before going to market. Is it a shirt, jeans, footwear etc.

· **Prepare**

Next, I need to check how much I am willing to spend and how to get that money.

· **Process**

Then I need to check how much I want from the data. Like if I am going to buy footwear what do I want - slippers / shoes / sandals etc.

· **Analyse**

You obviously won't buy things which are out of trend, also you need to check does the jeans which you have and the colour of t-shirt you want to buy, will it make a good combination.

· **Share**

Now you communicate your idea to the shopkeeper to find the best suitable fit for you.

· **Act**

Then you finally buy it!

Your Task:

Your task is to give the example(s) of such a real-life situation where we use Data Analytics and link it with the data analytics process. You can prepare a PPT/PDF on a real-life scenario explaining it with the above process (Plan, Prepare, Process, Analyse, Share, Act) and submit it as part of this task.

Plan : Decided to Vacation Due to Summer heat. Destination should beat the Summer

Heat like a hill station.

Prepare : Check the bank for cash and decide how much you can spend on the trip.

Whether Use credit or Saved Cash. Choose a form of travel according to budget

Process : Research on Places like Ooty , Coorg , Chikmagalur .

Choose a Place - Ooty

Collect flight prices & local prices

Prepare a Plan on how to spend time accommodation food etc in ooty

Analyze : Analyze the travel route and which Takes less time From bangalore to ooty.

Analyze the prices of hotels or resorts

Share : Share the trip plan to friends and ask them more ideas to experience the place

and modify the plan.

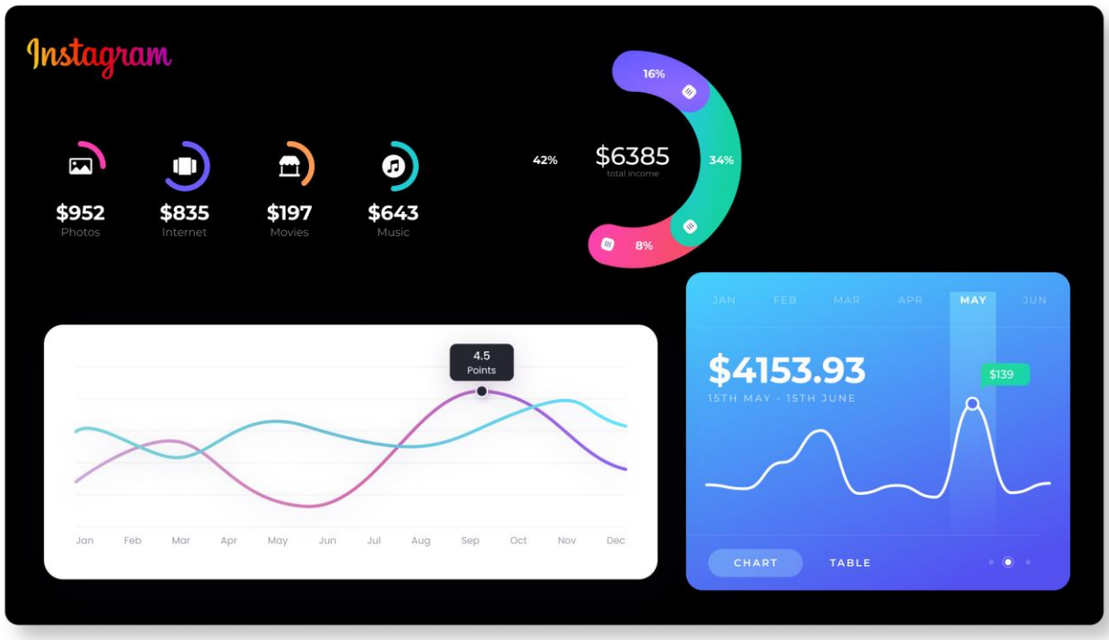
Act : Book Flight , Accommodation and have a list of all Activities which i want to

experience. Pack all the essential needs! And Leave for a Nice Vacation

This example shows how the data analytics process can be applied to real life business problems and how it can help organizations make data-driven decisions.

Module - 2

INSTAGRAM USER ANALYTICS



Project Description :

The aim of this project is to perform analysis on Instagram user data and gain insights into the behaviour of Instagram users. SQL was used as a primary tool for data analysis and retrieval. The findings from the analysis provide valuable insights into user activity, preferences and demographics on the popular social media platform. This report presents the methodology and results of the analysis and discusses the implications of the findings for businesses and marketers."

Approach:

• Database Creation: Use SQL to create a database that can store the data and support the queries you need to perform the analysis. • Data Loading: Load the data into the database.

• Data Analysis: Use SQL queries to perform the analysis and retrieve the insights you want to extract.

Tech-Stack Used :

MySQL and Adobe PDF

Insights : Below I have provided SQL queries and their respective outputs and with insights/outcomes for each of the 7 questions.

A) Marketing Analysis:

Loyal User Reward

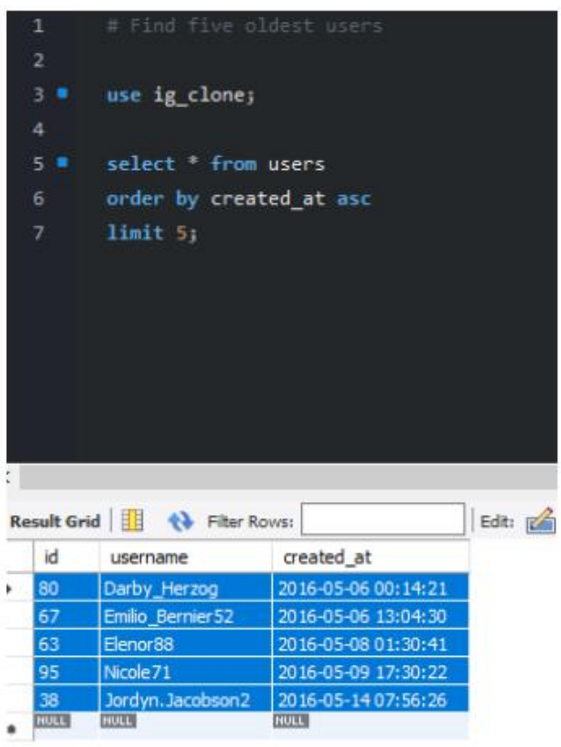
Query :

USE ig_clone;

SELECT * FROM users

ORDER BY created_at asc

LIMIT 5;



```
# Find five oldest users

use ig_clone;

select * from users
order by created_at asc
limit 5;
```

Result Grid | Filter Rows: | Edit:

id	username	created_at
80	Darby_Herzog	2016-05-06 00:14:21
67	Emilio_Bernier52	2016-05-06 13:04:30
63	Elenor88	2016-05-08 01:30:41
95	Nicole71	2016-05-09 17:30:22
38	Jordyn.Jacobson2	2016-05-14 07:56:26
NULL	NULL	NULL

2. Inactive User Engagement:

The team wants to encourage inactive users to start posting by sending them promotional emails.

Your Task: Identify users who have never posted a single photo on Instagram.

Query :

use ig_clone;

SELECT u.id, u.username

FROM users u

LEFT JOIN photos p ON u.id = p.user_id

WHERE p.user_id IS NULL;

```
use ig_clone;

SELECT u.id, u.username
FROM users u
LEFT JOIN photos p ON u.id = p.user_id
WHERE p.user_id IS  NULL;
```

Result Grid | Filter Rows: | Export:

id	username
5	Aniya_Hackett
7	Kasandra_Homenick
14	Jaclyn81
21	Rocio33
24	Maxwell.Halvorson
25	Tierra.Trantow
34	Pearl7
36	Ollie_Ledner37
41	Mckenna17
45	David.Osinski47
49	Morgan.Kassulke
53	Linnea59
54	Duane60
57	Julien_Schmidt
66	Mike.Auer39
68	Franco_Keebler64
71	Nia_Haag
74	Hulda.Macejkovic
75	Leslie67
76	Janelle.Nikolaus81
80	Darby_Herzog
81	Esther.Zulauf61
83	Bartholome.Bernhard
89	Jessyca_West
90	Esmeralda.Mraz57
91	Bethany20

3.Contest Winner Declaration:

The team has organized a contest where the user with the most likes on a single photo wins.

Your Task: Determine the winner of the contest and provide their details to the team

Query :

use ig_clone;

SELECT users.id, users.username, photos.id AS photo_id, COUNT(likes.user_id) AS likes_count

FROM photos

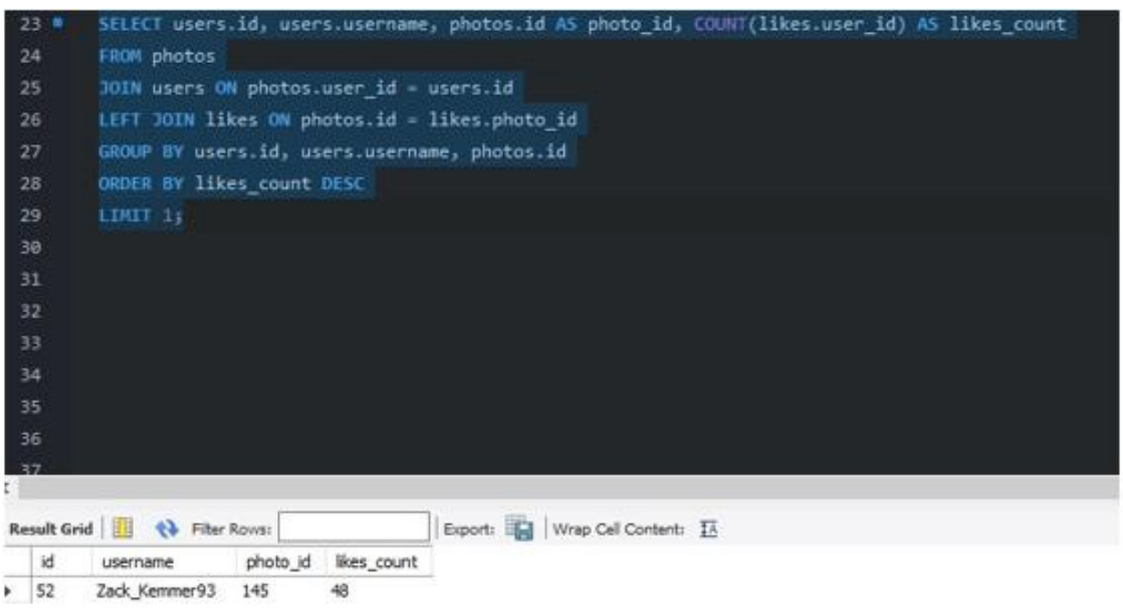
JOIN users ON photos.user_id = users.id

LEFT JOIN likes ON photos.id = likes.photo_id

GROUP BY users.id, users.username, photos.id

ORDER BY likes_count DESC

LIMIT 1;



4.Hashtag Research:

A partner brand wants to know the most popular hashtags to use in their

posts to reach the most people.

Your Task: Identify and suggest the top five most commonly used hashtags on the platform.

Query :

use ig_clone;

SELECT tags.tag_name, COUNT(*) AS tag_count

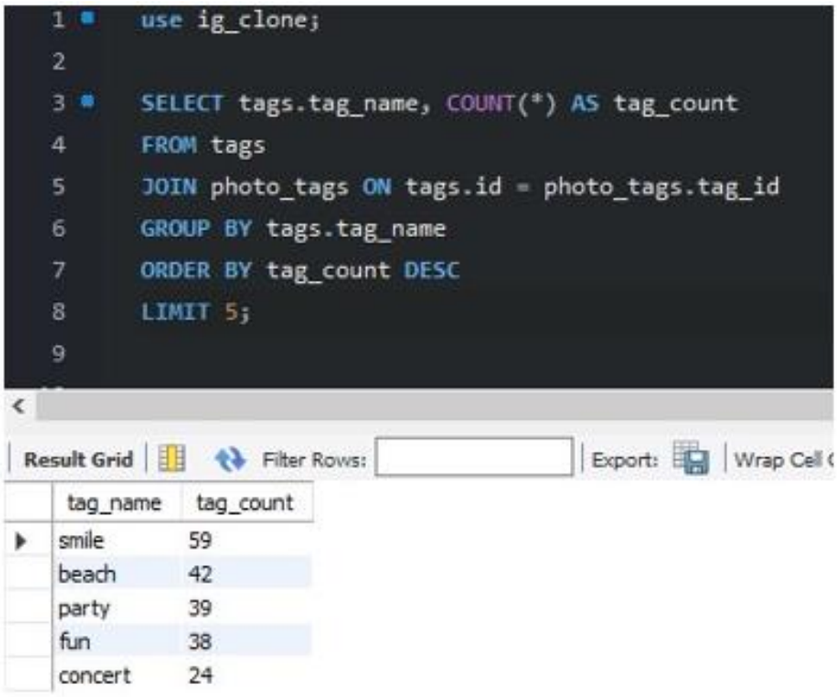
FROM tags

JOIN photo_tags ON tags.id = photo_tags.tag_id

GROUP BY tags.tag_name

ORDER BY tag_count DESC

LIMIT 5;



5.Ad Campaign Launch:

The team wants to know the best day of the week to launch ads.

Your Task: Determine the day of the week when most users register on Instagram. Provide

insights on when to schedule an ad campaign.

Query:

use ig_clone;

select dayname(created_at) As Day , Count(*) as Total

from users

group by day

order by total desc

limit 1;

```
select dayname(created_at) As Day , Count(*) as Total
from users
group by day
order by total desc
limit 1;
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

Day	Total
Thursday	16

B) Investor Metrics:

1. User Engagement: Investors want to know if users are still active and posting on

Instagram or if they are making fewer posts.

Your Task: Calculate the average number of posts per user on Instagram. Also,

provide the total number of photos on Instagram divided by the total number of

users.

Query :

use ig_clone;

select

(Selectcount(*) from photos) / (select count(*) from users) as average;

select user_id , count(*) as num_likes

from likes

group by user_id

having num_likes = (Select count(*) from photos);

select users.username , count(*) as num_likes

from users

join likes on id = likes.user_id

group by id

having num_likes = (Select count(*) from photos);

```
select
(Select count(*) from photos) / (select count(*) from users) as average;

select user_id , count(*) as num_likes
from likes
group by user_id
having num_likes = (Select  count(*) from photos);

select users.username , count(*) as num_likes
from users
join likes on id = likes.user_id
group by id
having num_likes = (Select  count(*) from photos);
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

username	num_likes
Aniya_Hackett	257
Jaclyn81	257
Rocio33	257
Maxwell.Halvorson	257
Ollie_Ledner37	257
Mckenna17	257
Duane60	257
Julien_Schmidt	257
Mike.Auer39	257
Nia_Haag	257
Leslie67	257
Janelle.Nikolaus81	257
Bethany20	257

Module - 3

Operation Analytics and Investigating Metric Spike

Description:

Project Agenda:

1. Project Description: In my role as the Data Analyst Lead at Microsoft, I work closely with teams across the company to analyse operational data, seeking opportunities for improvement and providing actionable insights. Our focus is on streamlining workflows, improving automation, and forecasting the company's growth trajectory. Additionally, we investigate fluctuations in daily engagement and sales metrics, pinpointing causes and offering strategies to address any fluctuations or declines.

2. Approach: In this project, we'll collect data from different company departments, ensuring its accuracy and organization in an SQL database. Through SQL queries and analysis, we'll uncover patterns and anomalies in data, focusing on metrics such as engagement, sales, and customer satisfaction. We'll investigate spikes in these metrics using historical data and collaborate with

teams for context. Insights from our analysis will guide process improvements, workflow optimization, and automation efforts.

3. Tech-Stack Used: Used MySQL Workbench 8.0 community server version 8.0.36

Case 1

Task:

1.Calculate the number of jobs reviewed per hour for each day in November 2020.

```
SELECT
    ds AS Dates,
    ROUND((COUNT(job_id) / SUM(time_spent)) * 3600) AS 'Jobs reviewied per hour per day'
FROM
    job_data
WHERE
    ds BETWEEN '2020-11-01' AND '2020-11-30'
GROUP BY ds;
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

Dates	Jobs reviewied per hour per day
2020-11-30	180
2020-11-29	180
2020-11-28	218
2020-11-27	35
2020-11-26	64
2020-11-25	80

Query:

SELECT

ds AS Dates,

ROUND((COUNT(job_id) / SUM(time_spent)) * 3600) AS 'Jobs reviewied per hour per

day'

FROM

job_data

WHERE

ds BETWEEN '2020-11-01' AND '2020-11-30'

GROUP BY ds;

Insights :

The number of jobs reviewed per hour per day in November 2020 varies,

with higher activity on some days and lower on others.

2. Calculate 7 day rolling average of throughput? For throughput, do you prefer daily

metric or 7- day rolling and why?

```
SELECT
    ROUND(COUNT(event) / SUM(time_spent), 2) AS 'Weekly Throughput'
FROM
    job_data;
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

Weekly Throughput
0.03

Query:

SELECT

ROUND(COUNT(event) / SUM(time_spent), 2) AS 'Weekly Throughput'

FROM

job_data;

```
SELECT
    ds AS dates,
    ROUND(COUNT(event) / SUM(time_spent), 2) AS 'daily throughput'
FROM
    job_data
GROUP BY ds
ORDER BY ds;
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

dates	daily throughput
2020-11-25	0.02
2020-11-26	0.02
2020-11-27	0.01
2020-11-28	0.06
2020-11-29	0.05
2020-11-30	0.05

Query:

```
SELECT

ds AS dates,

ROUND(COUNT(event) / SUM(time_spent), 2) AS 'daily throughput'

FROM

job_data

GROUP BY ds

ORDER BY ds;
```

Insights :

The 7-day rolling average of throughput provides a levelled view of the data, allowing you to observe trends over time without being affected by daily variations.

Continue using the 7-day rolling average for throughput analysis, as it provides a more steady representation of performance trends. This can help in identifying long-term patterns and making more informed decisions.

3.Calculate the percentage share of each language in the last 30 days?

```sql
SELECT
    language AS Languages,
    ROUND(100 * COUNT(*) / sub.total, 2) AS Percentage
FROM
    job_data
CROSS JOIN
    (SELECT COUNT(*) AS total FROM job_data) sub
GROUP BY
    language, sub.total;
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

Languages	Percentage
English	12.50
Arabic	12.50
Persian	37.50
Hindi	12.50
French	12.50
Italian	12.50

Query:

SELECT

language AS Languages,

ROUND(100 * COUNT(*) / sub.total, 2) AS Percentage

FROM

job_data

CROSS JOIN

(SELECT COUNT(*) AS total FROM job_data) sub

GROUP BY

language, sub.total;

Insights :

The language distribution in the last 30 days is relatively balanced, with Persian having

the highest share. Consider investing in language-specific content or features to

enhance user engagement in languages with lower shares

4. Display Duplicates

```
#Let's say you see some duplicate rows in the data. How will you display duplicates from the table?

SELECT actor_id, COUNT(*) AS Duplicates FROM job_data

GROUP BY actor_id HAVING Duplicates > 1;
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

Languages	Percentage
English	12.50
Arabic	12.50
Persian	37.50
Hindi	12.50
French	12.50
Italian	12.50

Query:

SELECT actor_id, COUNT(*) AS Duplicates FROM job_data

GROUP BY actor_id HAVING Duplicates > 1;

Insights:

There is one duplicate row in the data based on the actor_id column. Implement data

validation mechanisms to prevent such duplicates in the future.

Case 2

1.Measure the activeness of users on a weekly basis.

```
SELECT
    EXTRACT(WEEK FROM occurred_at) AS weeks,
    COUNT(DISTINCT user_id) AS no_of_users
FROM
    events
WHERE
    event_type = 'engagement'
GROUP BY weeks
ORDER BY weeks;
```

Result Grid | Filter Rows: | Export: | Wr

weeks	no_of_users
17	663
18	1068
19	1113
20	1154
21	1121
22	1186
23	1232
24	1275
25	1264
26	1302
27	1372
28	1365
29	1376
30	1467
31	1299
32	1225
33	1225
34	1204
35	104

Query :

SELECT

EXTRACT(WEEK FROM occurred_at) AS weeks,

COUNT(DISTINCT user_id) AS no_of_users

FROM

events

WHERE

event_type = 'engagement'

GROUP BY weeks

ORDER BY weeks;

Insights

User engagement seems to have peaked around week 30 and has shown some variations over the observed period. Look for patterns related to content updates, marketing campaigns, or any external

events that might have influenced user behaviour. Use these insights to plan future

engagement strategies

2. Analyse the growth of users over time for a product.

```sql
SELECT
    week_num,
    year_num,
    SUM(active_users) OVER (
        ORDER BY week_num, year_num
        ROWS BETWEEN UNBOUNDED PRECEDING AND CURRENT ROW
    ) AS cumulative_sum
FROM (
    SELECT
        EXTRACT(WEEK FROM activated_at) AS week_num,
        EXTRACT(YEAR FROM activated_at) AS year_num,
        COUNT(DISTINCT user_id) AS active_users
    FROM
        users
    WHERE
        state LIKE '%active%'
    GROUP BY
        year_num, week_num
    ORDER BY
        year_num, week_num
) AS tab;
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

week_num	year_num	cumulative_sum
0	2013	23
0	2014	106
1	2013	136
1	2014	262
2	2013	310
2	2014	419
3	2013	455
3	2014	568
4	2013	598
4	2014	728
5	2013	776
5	2014	909

This is just a sample output.. the actual result contains 90 rows

Query :

```
SELECT
week_num,
year_num,
SUM(active_users) OVER (
ORDER BY week_num, year_num
ROWS BETWEEN UNBOUNDED PRECEDING AND CURRENT ROW
) AS cumulative_sum
FROM (
SELECT
EXTRACT(WEEK FROM activated_at) AS week_num,
EXTRACT(YEAR FROM activated_at) AS year_num,
COUNT(DISTINCT user_id) AS active_users
FROM
users
WHERE
state LIKE '%active%'
GROUP BY
year_num, week_num
ORDER BY
year_num, week_num
) AS tab;
```

Insights :

User growth has generally been positive over time, with some variations.

When the number of users is going up, it's smart to figure out why. Look at times when

more people are joining and see if it's because of things like making the product better,

doing more marketing, or changes in what people want. Understanding this helps keep

the growth going and maybe make it even faster.

3. Calculate the weekly retention of users-signup cohort

```sql
SELECT
    EXTRACT(WEEK FROM occurred_at) AS weeks,
    COUNT(DISTINCT user_id) AS no_of_users
FROM
    events
WHERE
    event_type = 'signup_flow'
        AND event_name = 'complete_signup'
GROUP BY weeks
ORDER BY weeks;
```

Result Grid | Filter Rows: | Export: | Wrap

weeks	no_of_users
17	72
18	163
19	185
20	176
21	183
22	196
23	196
24	229
25	207
26	201
27	222
28	215
29	221
30	238
31	193
32	245
33	261
34	259
35	18

Query :

```
SELECT
EXTRACT(WEEK FROM occurred_at) AS weeks,
COUNT(DISTINCT user_id) AS no_of_users
FROM
events
WHERE
event_type = 'signup_flow'
AND event_name = 'complete_signup'
GROUP BY weeks
ORDER BY weeks;
```

Insights:

Weekly user retention shows a slow decline over time.

Concentrate on improving user retention strategies. Identify key touchpoints

in the user journey where users might be dropping off and work on

enhancing user experience, engagement, and value during those stages.

4. Measure the activeness of users on a weekly basis per device.

```
SELECT
    device,
    EXTRACT(WEEK FROM occurred_at) AS weeks,
    COUNT(DISTINCT user_id) AS no_of_users
FROM
    events
WHERE
    event_type = 'engagement'
GROUP BY device , weeks
ORDER BY weeks;
```

Result Grid | Filter Rows: | Export: | Wrap

device	weeks	no_of_users
acer aspire desktop	17	9
acer aspire notebook	17	20
amazon fire phone	17	4
asus chromebook	17	21
dell inspiron desktop	17	18
dell inspiron notebook	17	46
hp pavilion desktop	17	14
htc one	17	16
ipad air	17	27
ipad mini	17	19
iphone 4s	17	21
iphone 5	17	65
iphone 5s	17	42
kindle fire	17	6
lenovo thinkpad	17	86
mac mini	17	6
macbook air	17	54
macbook pro	17	143
nexus 10	17	16
nexus 5	17	40
nexus 7	17	18
nokia lumia 635	17	17

Result 4

This is just a sample output.. the result contains 491 rows.

Query :

SELECT

device,

EXTRACT(WEEK FROM occurred_at) AS weeks,

COUNT(DISTINCT user_id) AS no_of_users

FROM

events

WHERE

event_type = 'engagement'

GROUP BY device , weeks

ORDER BY weeks;

Insights :

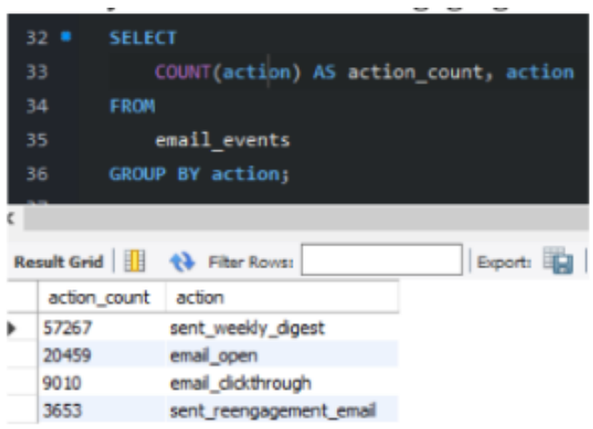
Engagement has variations across different devices and weeks. Some devices

consistently show higher engagement when compared to others.

Optimize the user experience on the devices which show lower engagement.

Consider high engagement devices and design strategies to reach high potential.

5. Analyse how users are engaging with the email service.



```
SELECT
    COUNT(action) AS action_count, action
FROM
    email_events
GROUP BY action;
```

Result Grid | Filter Rows: | Export:

action_count	action
57267	sent_weekly_digest
20459	email_open
9010	email_clickthrough
3653	sent_reengagement_email

Query :

SELECT

COUNT(action) AS action_count, action

FROM

email_events

GROUP BY action;

This query gives us the action count of every action corresponding in the table.

```sql
SELECT
    (SUM(CASE
        WHEN email_category = 'email_opened' THEN 1
        ELSE 0
    END) / SUM(CASE
        WHEN email_category = 'email_sent' THEN 1
        ELSE 0
    END)) * 100 AS open_rate,
    (SUM(CASE
        WHEN email_category = 'email_clickthrough' THEN 1
        ELSE 0
    END) / SUM(CASE
        WHEN email_category = 'email_sent' THEN 1
        ELSE 0
    END)) * 100 AS click_rate
FROM
    (SELECT
        *,
            CASE
                WHEN action IN ('sent_weekly_digest' , 'sent_reengagement_email') THEN ('email_sent')
                WHEN action IN ('email_open') THEN ('email_opened')
                WHEN action IN ('email_clickthrough') THEN ('email_clickthrough')
            END AS email_category
    FROM
        email_events) AS alias;
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

open_rate	click_rate
33.5834	14.7899

This query gives us email open rate and click rate.

Query:

SELECT

(SUM(CASE

WHEN email_category = 'email_opened' THEN 1

ELSE 0

END) / SUM(CASE

WHEN email_category = 'email_sent' THEN 1

ELSE 0

END)) * 100 AS open_rate,

(SUM(CASE

WHEN email_category = 'email_clickthrough' THEN 1

```
ELSE 0
END) / SUM(CASE

WHEN email_category = 'email_sent' THEN 1
ELSE 0
END)) * 100 AS click_rate
FROM
(SELECT
*,
CASE
WHEN action IN ('sent_weekly_digest' , 'sent_reengagement_email')
THEN ('email_sent')
WHEN action IN ('email_open') THEN ('email_opened')
WHEN action IN ('email_clickthrough') THEN ('email_clickthrough')
END AS email_category
FROM
email_events) AS alias;
```

Insights :

The email engagement metrics show an open rate of approximately 33.58% and a click rate of about 14.79%.

Result

This project deepened my grasp of operational analytics and data integration

Through SQL analysis, I refined my skills in querying complex datasets. I have learned how to effectively merge and normalize various datasets, ensuring accurate and reliable analysis.

Collaboration with diverse teams honed my communication, teamwork, and ability to translate data insights into actionable recommendations for process improvement.

Hiring Process Analytics

Project Agenda

1. Project description: The aim of this project is to analyse the hiring trends of a multinational company (MNC) using a dataset provided by the company. As the lead Data Analyst at a prestigious organization like Google, the task is to extract significant insights from the data and provide a detailed report to the hiring department. The dataset includes the details about people who registered for a particular post in a department of this company.

2. Approach: The project follows a methodical approach to Exploratory Data Analysis (EDA), including understanding the data columns and their content, checking for missing data, clubbing columns with multiple categories for comprehensive analysis, identifying and handling outliers, and creating a data summary. Using statistical knowledge and Excel formulas, the project aims to draw meaningful conclusions about the company's hiring trends. The detailed report will provide actionable insights to the hiring department, aiding in decision making and improving the overall hiring process.

3. Tech-stack used: The primary tech stack for this project would involve using Microsoft Excel. Excel offers a wide range of functions

and tools for data analysis and manipulation, making it ideal for tasks like exploratory data analysis and drawing insights from the dataset. Excel's features, such as formulas, functions, pivot tables, and charts, will be utilized for data cleaning, calculating statistics, identifying trends, and creating visual representations of the hiring data, providing a user-friendly interface for analysis

Task 1 :

Determine the gender distribution of hires. How many males and females have been hired by the company?

Insight:

The company hired more males (2563) compared to females (1856).

Company may conduct a gender diversity review to ensure unbiassed hiring practices.

Task 2 :

What is the average salary offered by this company? Use Excel functions to calculate this.

D
Avg salary of employees
49752.90

Insight:

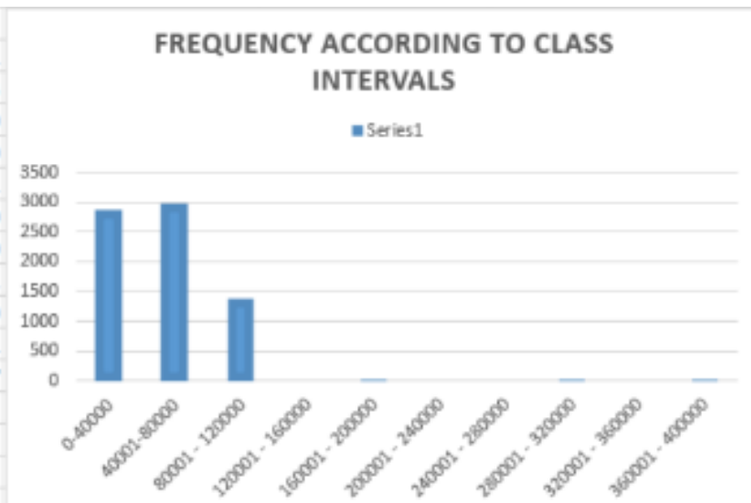
The average salary offered in the company is $49,752.90.

The company can measure whether this average salary aligns with industry standards and competitors. If the average is significantly lower, they may need to re-evaluate their compensation packages to attract and retain top talent

Task 3 :

Create class intervals for the salaries in the company. This will help you understand the salary distribution.

Class Intervals	Frequency
0-40000	2831
40001-80000	2963
80001 - 120000	1370
120001 - 160000	0
160001 - 200000	1
200001 - 240000	0
240001 - 280000	0
280001 - 320000	1
320001 - 360000	0
360001 - 400000	1
	7167



Insight:

The majority of employees fall within the salary range of $0-$80,000.

The company can use this data to analyse the distribution of salaries and

consider whether adjustments are needed to ensure fair and competitive

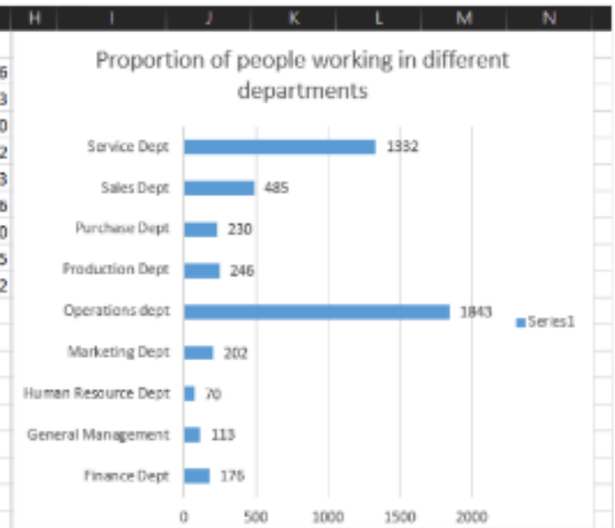
compensation, especially for roles with no representation in certain salary

ranges.

Task 4:

Use a pie chart, bar graph, or any other suitable visualization to show the proportion of people working in different departments.

Proportion of people working in different deptartments	
Finance Dept	176
General Management	113
Human Resource Dept	70
Marketing Dept	202
Operations dept	1843
Production Dept	246
Purchase Dept	230
Sales Dept	485
Service Dept	1332



Insight:

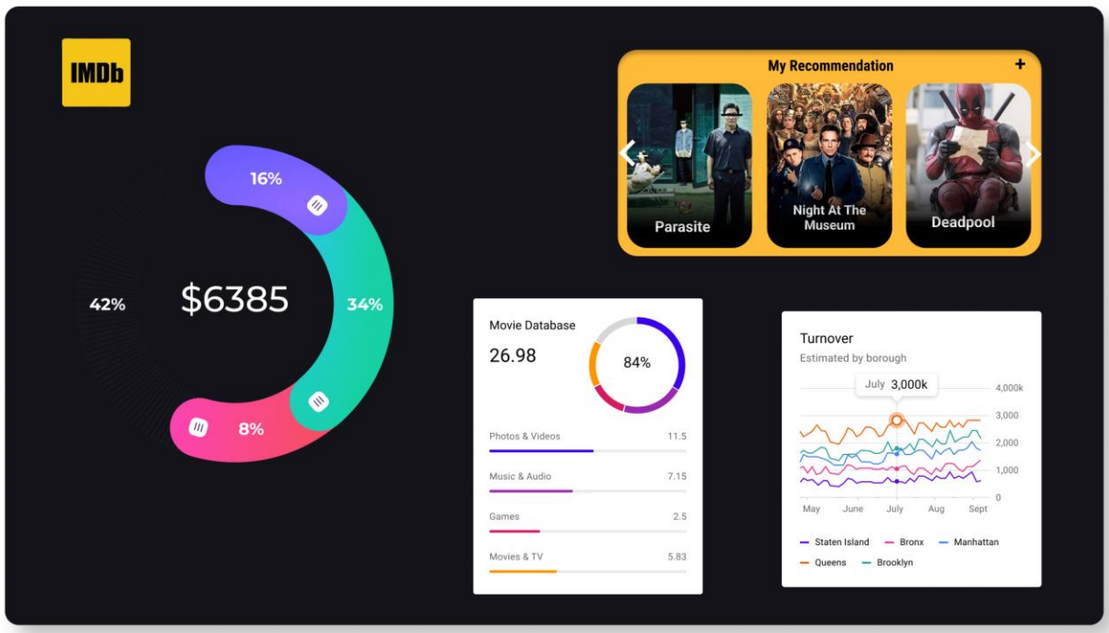
The majority of employees work in the "Operations" and "Service" departments. The company can calculate departmental distribution and consider to strengthen other departments by adjusting their recruitment efforts or offer incentives to attract more talent in those areas.

Result

While working on this project, I have achieved several accomplishments. I successfully analysed the hiring data, identified the number of males and females hired, as well as calculated the average salary offered. Additionally, I created class intervals for salary ranges and visualized the data through charts and graphs to showcase departmental proportions and post tiers. This project has enhanced my skills in data analysis,

statistical calculations, and data visualization techniques. It has also provided me with valuable experience in deriving insights from real-world datasets, which will further strengthen my capabilities as a data analyst.

IMDB Movie Analysis



PROJECT DESCRIPTION:

The IMDB Movie Analysis project aims to explore and analyse a comprehensive dataset of movies available on the IMDB platform. This dataset contains essential information about movies, including director names, movie titles, duration, genre, budget, gross earnings, IMDb ratings, and more. Through in-depth data analysis using Excel, Data Visualization and Statistics techniques this project seeks to extract valuable insights and trends that contribute to a movie's success.

My Approach:

I have gone through the dataset and understood all the given columns. Then I have observed that there are total of 28 columns and 5031 rows. This dataset also contains unwanted columns. Null Values and blank rows. So, I decided to clean the dataset for analysis:

1) I deleted the columns which is not required for our analysis and left out with only 9 columns which are Director’s Name, Duration ,

Movie Title , genre, Budget , gross , imdb_rating, Language and country.

2) Then selected all the blank cells using go to special, select all and delete entire rows with blank cells.

3) Search for duplicate data and deleted all the duplicates,

4) Removed the Â from all the movies names using find and replace.

Now we're left with 3854 rows and 9 Columns.

Tech Stack:

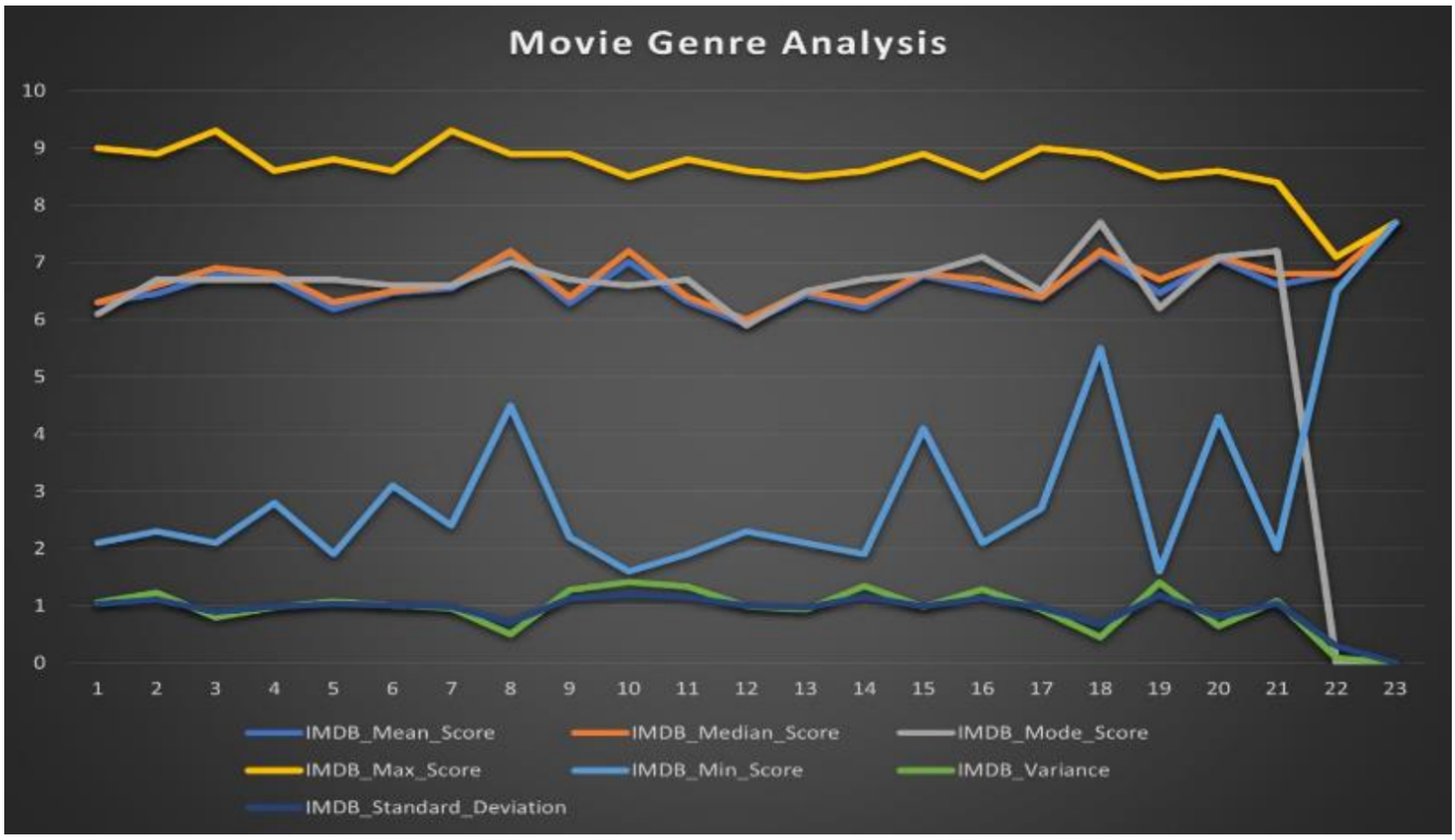
For this project I used Microsoft excel 365 to run functions and get answers for the above questions and also graphs.

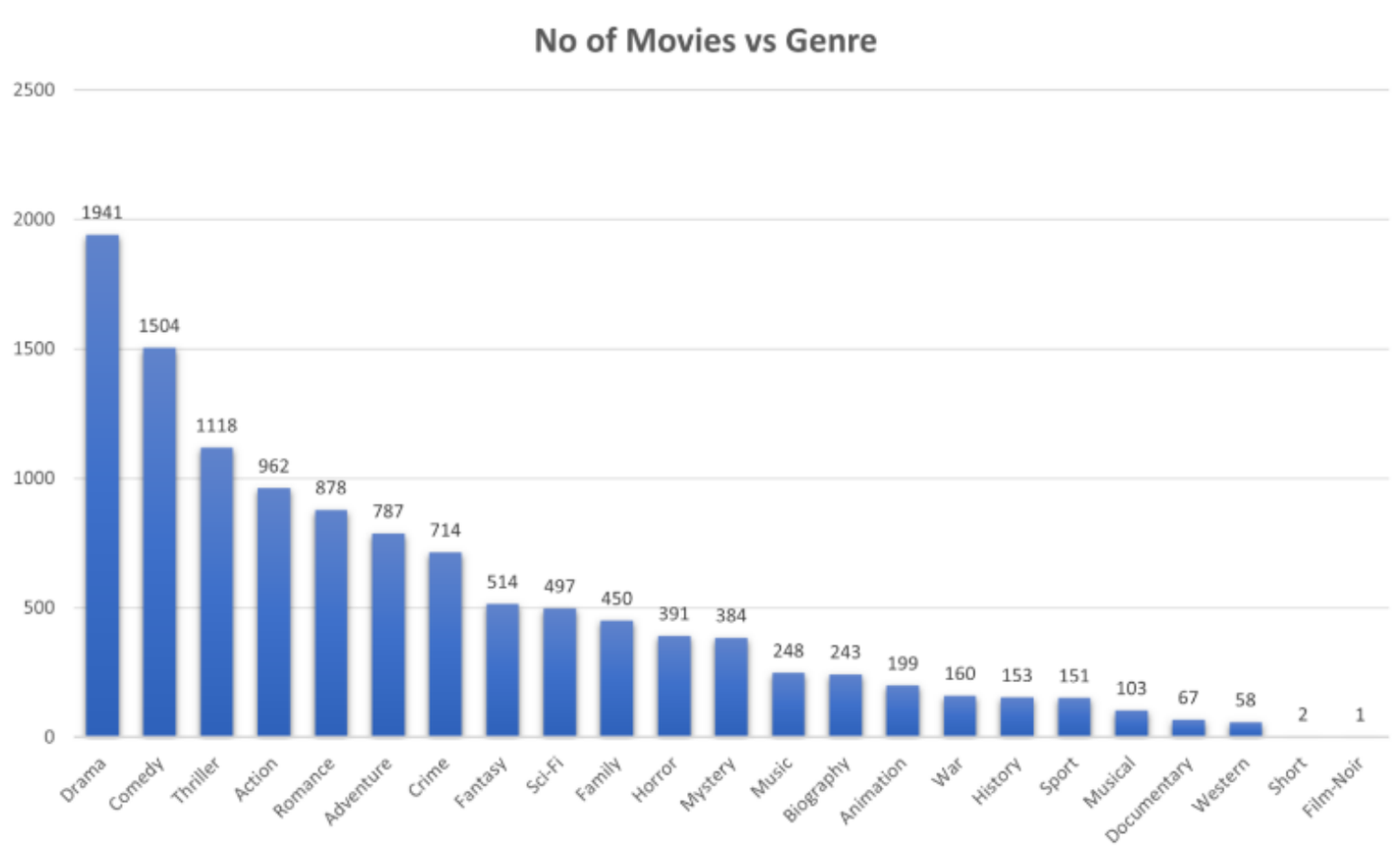
Insights :

A. Movie Genre Analysis: Analyse the distribution of movie genres and their effect on the IMDB score.

● Task: Determine the most common genres of movies in the dataset. Then, for each genre, calculate descriptive statistics (mean, median, mode, range, variance, standard deviation) of the IMDB scores.

Genres	No of Movies	IMDB_Mean_Score	IMDB_Median_Score	IMDB_Mode_Score	IMDB_Max_Score	IMDB_Min_Score	IMDB_Variance	IMDB_Standard_Deviation
Action	962	6.290748441	6.3	6.1	9	2.1	1.062648296	1.030848338
Adventure	787	6.458322745	6.6	6.7	8.9	2.3	1.226420567	1.107438742
Drama	1941	6.786192684	6.9	6.7	9.3	2.1	0.793647586	0.890869006
Animation	199	6.700502513	6.8	6.7	8.6	2.8	0.972411808	0.98610943
Comedy	1504	6.184441489	6.3	6.7	8.8	1.9	1.075981337	1.037295202
Mystery	384	6.472916667	6.5	6.6	8.6	3.1	1.008641493	1.004311452
Crime	714	6.544817927	6.6	6.6	9.3	2.4	0.959363902	0.979471236
Biography	243	7.141563786	7.2	7	8.9	4.5	0.500618131	0.707543731
Fantasy	514	6.290077821	6.4	6.7	8.9	2.2	1.27412334	1.128770721
Documentary	67	7.011940299	7.2	6.6	8.5	1.6	1.418364892	1.190951255
Sci-Fi	497	6.322736419	6.4	6.7	8.8	1.9	1.333869373	1.154932627
Horror	391	5.926086957	6	5.9	8.6	2.3	0.991646836	0.99581466
Romance	878	6.42881549	6.5	6.5	8.5	2.1	0.933600191	0.966229885
Family	450	6.210444444	6.3	6.7	8.6	1.9	1.348579802	1.161283687
Western	58	6.765517241	6.8	6.8	8.9	4.1	0.979845422	0.989871417
Musical	103	6.559223301	6.7	7.1	8.5	2.1	1.289211047	1.135434299
Thriller	1118	6.377370304	6.4	6.5	9	2.7	0.932341206	0.965578172
History	153	7.136601307	7.2	7.7	8.9	5.5	0.449510018	0.670455082
Music	248	6.460483871	6.7	6.2	8.5	1.6	1.40553525	1.185552719
War	160	7.05375	7.1	7.1	8.6	4.3	0.648735937	0.805441455
Sport	151	6.603311258	6.8	7.2	8.4	2	1.081379764	1.039894112
Short	2	6.8	6.8	0	7.1	6.5	0.09	0.3
Film-Noir	1	7.7	7.7	0	7.7	7.7	0	0



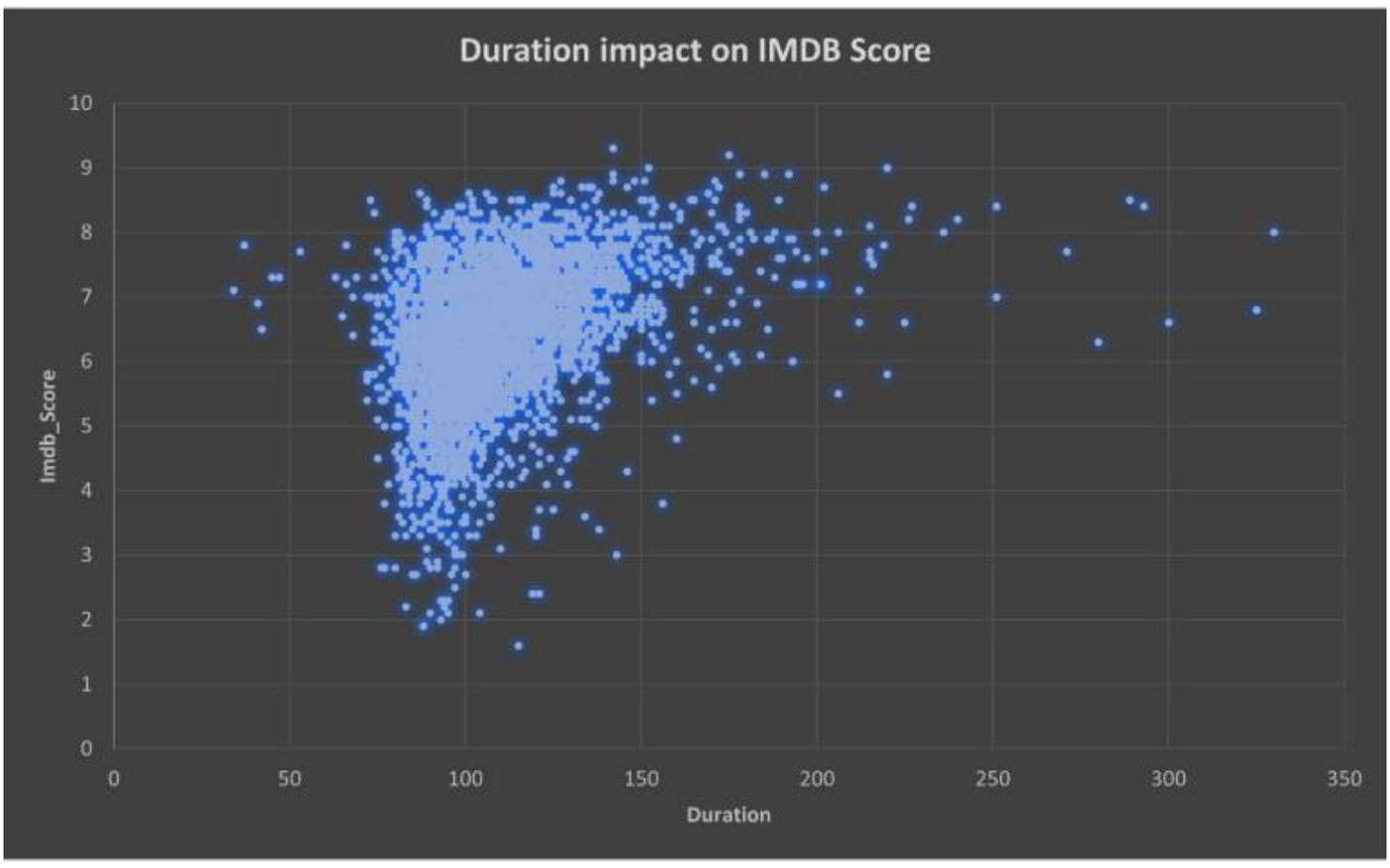


The Most common movie genres from the dataset are Drama, Comedy, Thriller and Action.

B. Movie Duration Analysis: Analyse the distribution of movie durations and its effect on the IMDB score.

● Task: Analyse the distribution of movie durations and identify the relationship between

Operations	Values
Mean	109.6647059
Median	105
Mode	101
Variance	516.8857876
Standard Deviation	22.73512233



movie duration and IMDB score.

The Average duration of a Movie is 109 minutes.

The movies have high average IMDB rating which are between 100-150 mins.

C. Language Analysis: Situation: Examine the distribution of movies based on their language.

● Task: Determine the most common languages used in movies and analyse their effect on the IMDB score using descriptive statistics.

Language	No Of Movies	imdb mean	imdb_median	imdb_mode	imdb_variance	imdb_standard_deviation
English	3729	6.41	6.5	6.7	1.123405775	1.05990838
French	37	7.29	7.2	7.2	0.306574142	0.553691378
Spanish	27	7.04	7.1	7.2	0.632839506	0.795512103
Dari	16	7.5	7.4	7.6	0.502773438	0.709065186
Mandarin	14	7.02	7.25	7.6	0.544540816	0.737930089
German	13	7.69	7.7	7.4	0.379171598	0.615769111
Japanese	12	7.63	7.8	0	0.741875	0.861321659
Hindi	10	6.76	7.05	0	1.1124	1.05470375
Cantonese	8	7.24	7.3	7.3	0.16984375	0.412121038
Italian	7	7.19	7	0	1.144081633	1.069617517

The Most common languages used in the movies are English, French, Spanish, Dari, Mandarin and German. I have also Observed that the languages Telugu and Persian have the highest average IMDB score

D. Director Analysis: Effect of directors on movie ratings.

● Task: Identify the top directors based on their average IMDB score and analyse their contribution to the success of movies using percentile calculations.

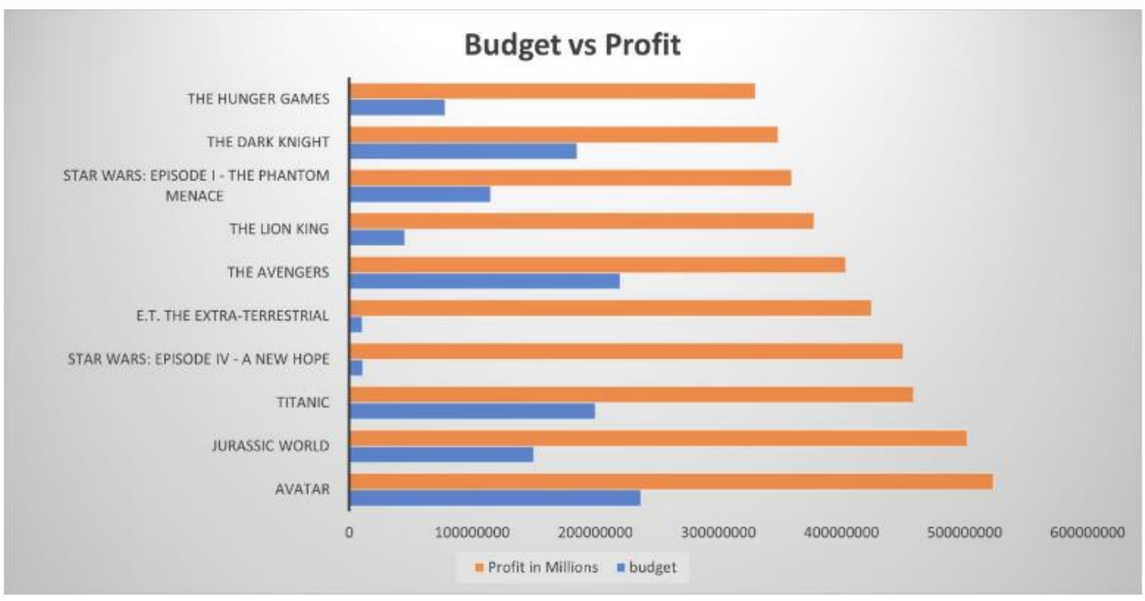
Director Name	No OF Movies	Average Imdb Score	Percentile
Tony Kaye	1	8.6	0.998
Charles Chaplin	1	8.6	0.998
Alfred Hitchcock	1	8.5	0.996
Ron Fricke	1	8.5	0.996
Damien Chazelle	1	8.5	0.996
Majid Majidi	1	8.5	0.996
Christopher Nolan	8	8.43	0.995
Sergio Leone	3	8.43	0.995
S.S. Rajamouli	1	8.4	0.993
Richard Marquand	1	8.4	0.993
Asghar Farhadi	1	8.4	0.993
Marius A. Markevicius	1	8.4	0.993
Lee Unkrich	1	8.3	0.991
Fritz Lang	1	8.3	0.991
Lenny Abrahamson	1	8.3	0.991
Billy Wilder	1	8.3	0.991
Pete Docter	3	8.23	0.99
Hayao Miyazaki	4	8.23	0.99

I have identified that Tony Kaye, Charles Chaplin, Alfred Hitchcock, Ron Fricke, Damien Chazelle, Majid Majidi, Sergio Leone, Christopher Nolan, SS Rajamouli, Richard Marquand, Asghar Farhadi and Marius A Markevicius are the top 10 directors with average IMDB score >=8.4.

E. Budget Analysis: Explore the relationship between movie budgets and their financial success.

• Task: Analyse the correlation between movie budgets and gross earnings, and identify the movies with the highest profit margin.

Movie Title	budget	Profit in Millions
Avatar	237000000	$52,35,05,847
Jurassic World	150000000	$50,21,77,271
Titanic	200000000	$45,86,72,302
Star Wars: Episode IV - A New Hope	11000000	$44,99,35,665
E.T. the Extra-Terrestrial	10500000	$42,44,49,459
The Avengers	220000000	$40,32,79,547
The Lion King	45000000	$37,77,83,777
Star Wars: Episode I - The Phantom Menace	115000000	$35,95,44,677
The Dark Knight	185000000	$34,83,16,061
The Hunger Games	78000000	$32,99,99,255

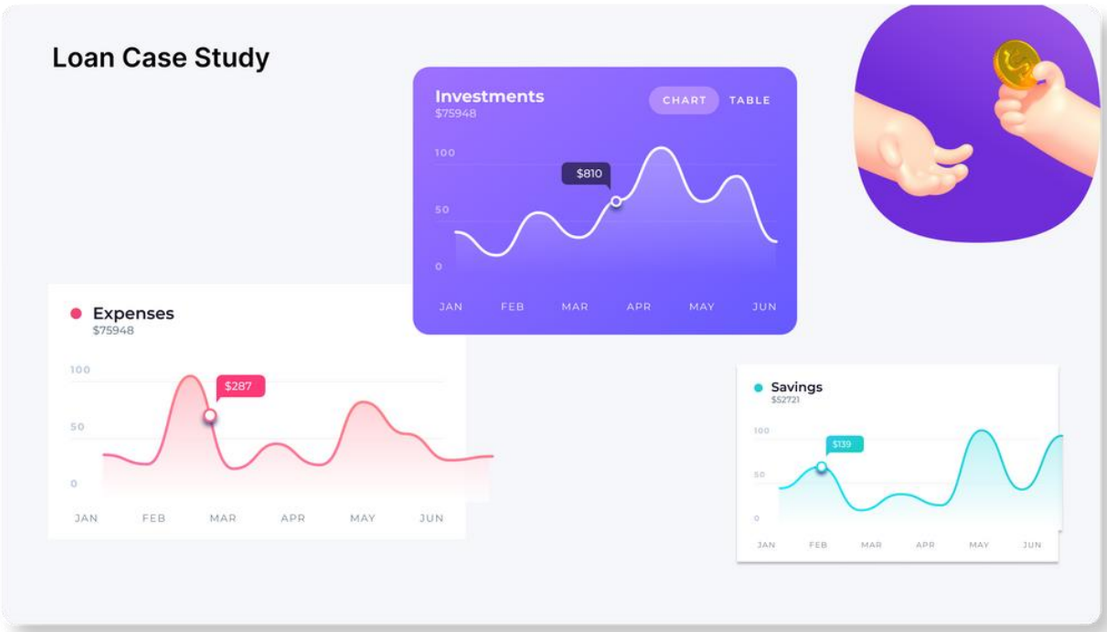


The Top-5 with highest profits are Avatar, Jurassic World, Titanic, Star Wars: Episode IV - A New Hope and E.T. The Extra-Terrestrial. The Correlation between budget and gross is positive.

RESULTS :

With the help of this project, I have gained valuable skill for data analysis using statistical knowledge and excel data visualization. Through this, I have learnt to apply my data analysis skills in solving real life problems

Bank Loan Case Study



Project Description:

The Bank Loan Case Study aims to identify Patterns which shows if the client has difficulty in paying their instalments which may be used for taking actions such as denying the loan, reducing the amount of loan, lending to risky applicants at a higher interest rate etc.

The dataset contains the information about the loan applicants at the time of applying for the loan. It contains two types of scenarios.

1. The client with repayment difficulties he/she had late payment more than x no of days on at least one of the first y instalment of the loan in our sample

2. All other cases when the payment is paid on time.

When a client applies for the loan, there are four types of decision that could be taken by client/Company

- **Approved**
- **Cancelled**

- **Refused**
- **Unused offer**

In this case study, we Use Exploratory Data Analysis to understand how consumer attributes and loan attributes influence the tendency of Default.

Tech Stack Used:

Microsoft Excel 2021 for Data analysis, Data Cleaning and Visualisations.

Adobe PDF For Presenting the Project.

Project Approach:

Understanding: The dataset set provided contains 3 files

Application Data : Contains current Applicant Data 50000 Rows and 122 Columns

Previous Application Data: Contains previous applicant data 50000 rows and 37 columns

Column description : Contains elaborated description of columns.

Data Analysis Tasks:

A. Identify Missing Data and Deal with it Appropriately: As a data analyst, you come across missing data in the loan application dataset. It is essential to handle missing data effectively to ensure the accuracy of the analysis.

Task: Identify the missing data in the dataset and decide on an appropriate method to deal with it using Excel built-in functions and features

Performed the first task which is finding out which columns has >40% null Values and removed all the columns with >40% null values.

Cleaned Dataset contains : 50000 rows and 73 columns...

- Used median to impute Amt_Annuity

• Used Mode for Name_type_suite

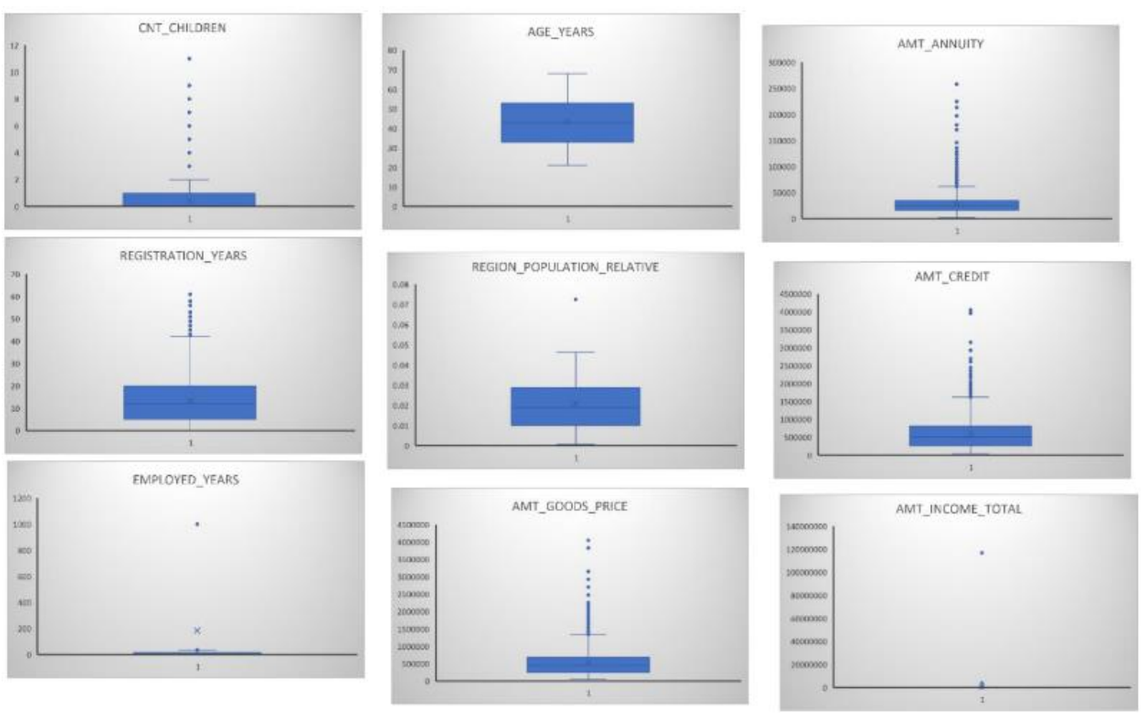
• Columns which had negative values for days Days_birth, Days_employed, Days_registration,, Used abs function and divided by 365 to convert it into years.

• Used Custom Impute Value “unknown” for occupation type as the null was high.

B. Identify Outliers in the Dataset: Outliers can significantly impact the analysis and distort the results. You need to identify outliers in the loan application dataset.

Task: Detect and identify outliers in the dataset using Excel statistical functions and features, focusing on numerical variables

Used Box and whisker chart to find outliers



Insights:

Count of children is 11 which is Our outlier of the dataset

Employed years had 1000 years which could be false info during data collection

Income which is more than 11Crores was an outlier

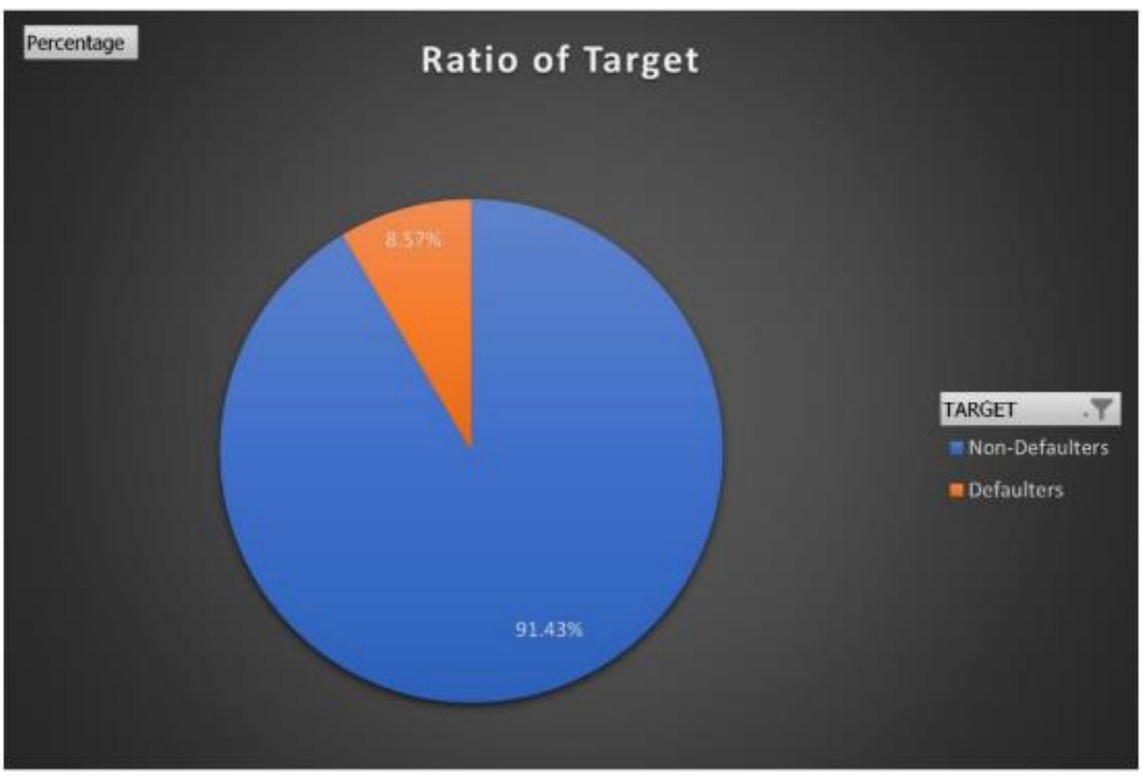
Data without outliers:

73 columns and 41074 rows.

C. Analyse Data Imbalance: Data imbalance can affect the accuracy of the analysis, especially for binary classification problems. Understanding the data distribution is crucial for building reliable models.

Task: Determine if there is data imbalance in the loan application dataset and calculate the ratio of data imbalance using Excel functions.

Data Imbalance Chart:



Insights:

We can see that we have

91.43% of Non-Defaulters

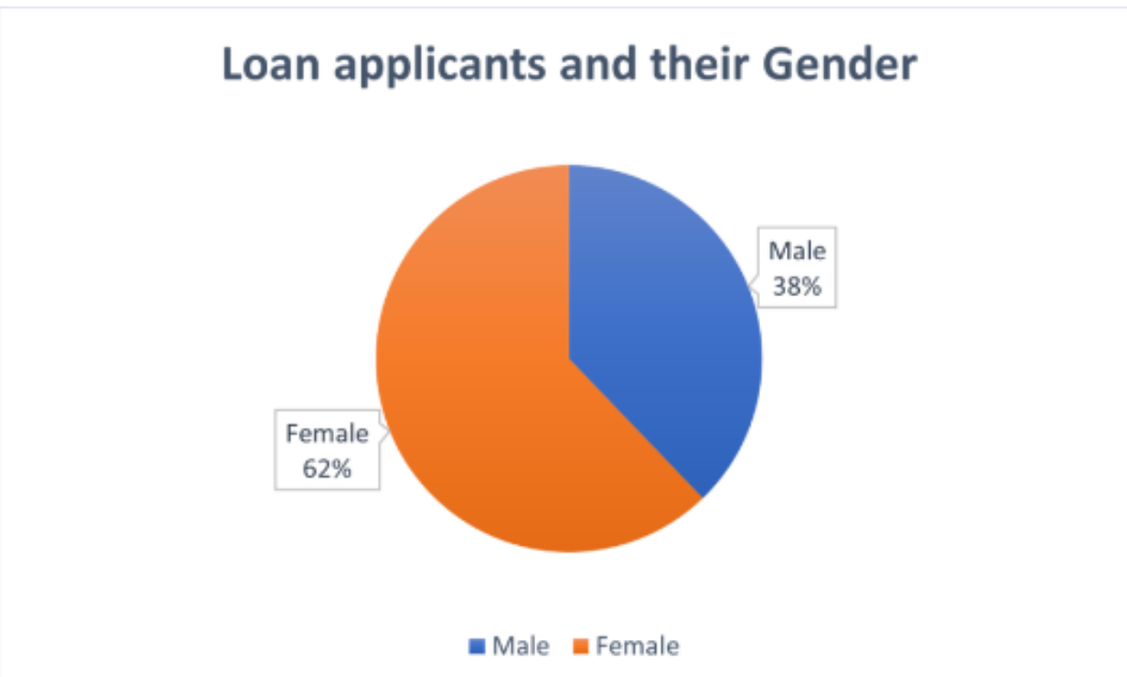
8.57% of Defaulters

Which clearly indicates Data is highly imbalanced.

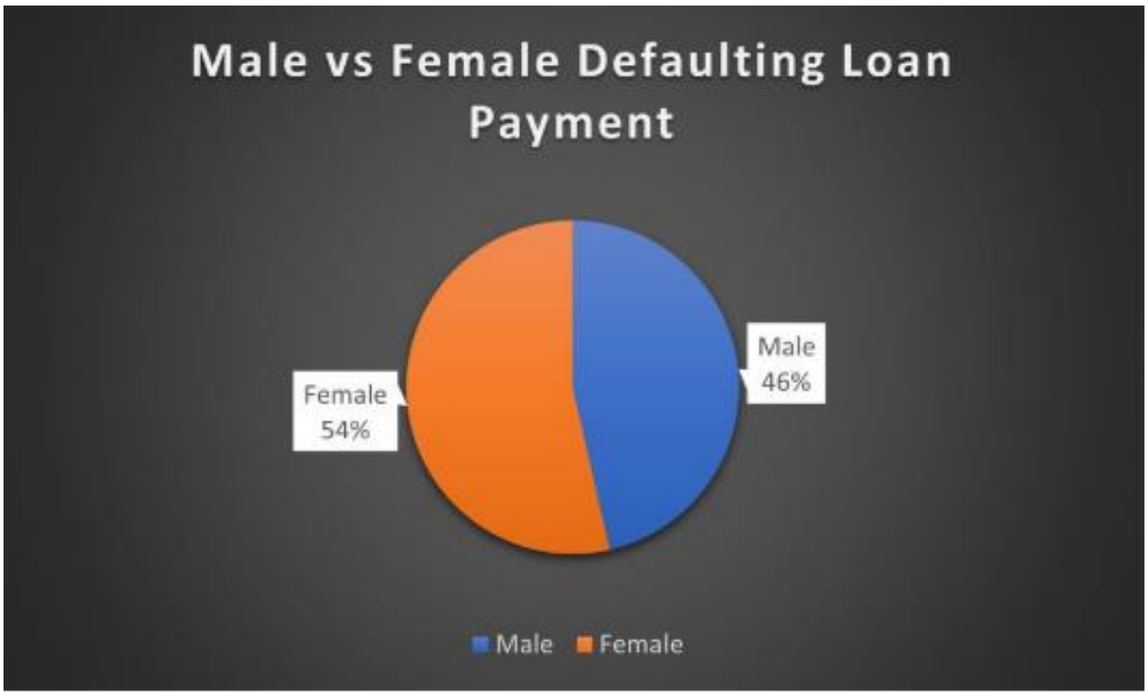
D. Perform Univariate, Segmented Univariate, and Bivariate Analysis: To gain insights into the driving factors ofloan default, it is important to conduct various analyses on consumer and loan attributes.

Task: Perform univariate analysis to understand the distribution of individual variables, segmented univariate analysis to compare variable distributions for different scenarios, and bivariate analysis to explore relationships between variables and the target variable using Excel functions and features.

1.Gender Ratio on Defaulting the Loan

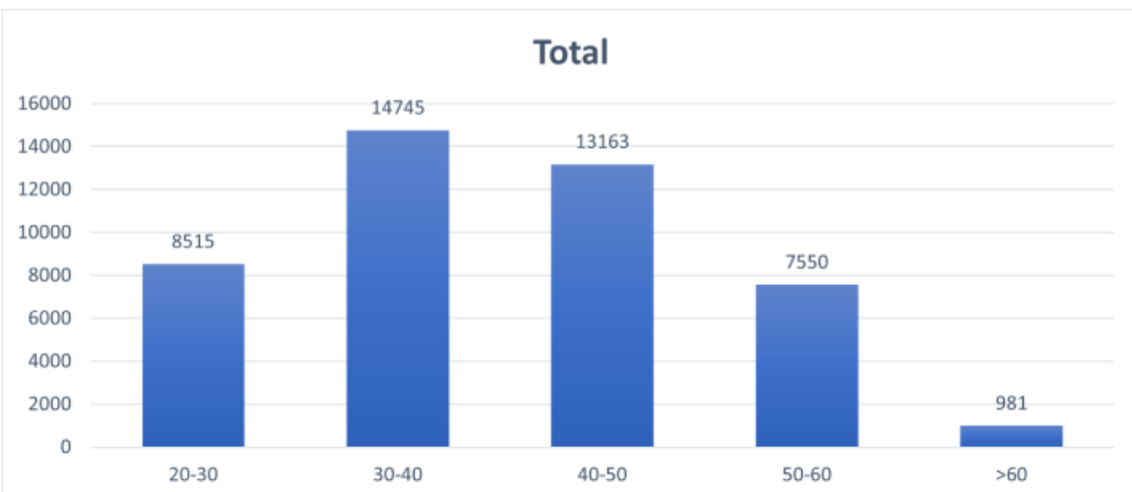


The number of Female Loan applicants is higher than Males.

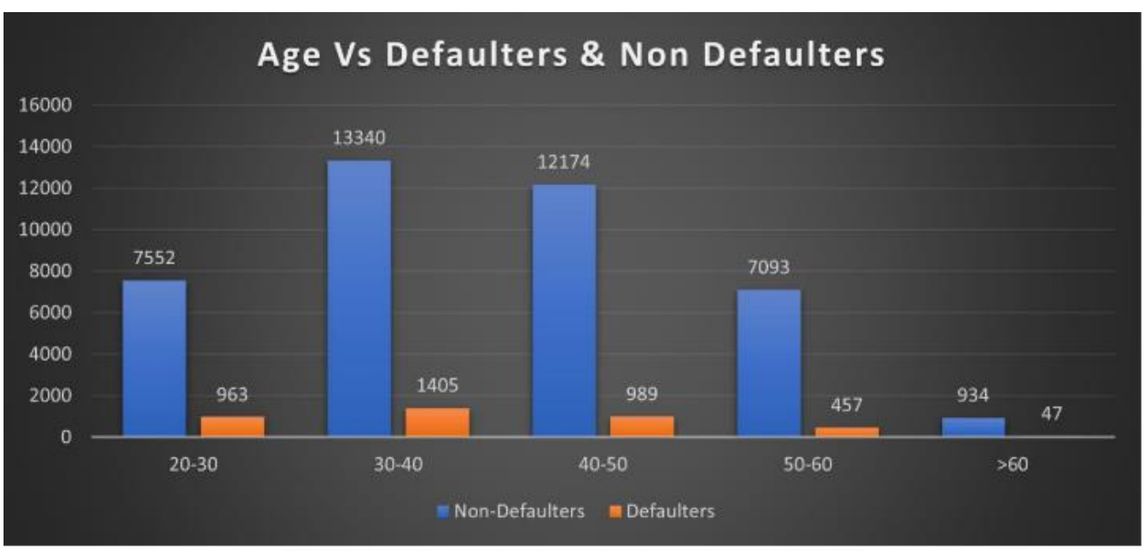


With this Graph we can see that most of the Defaulters are female.

2. Age Univariate and Segmented Univariate

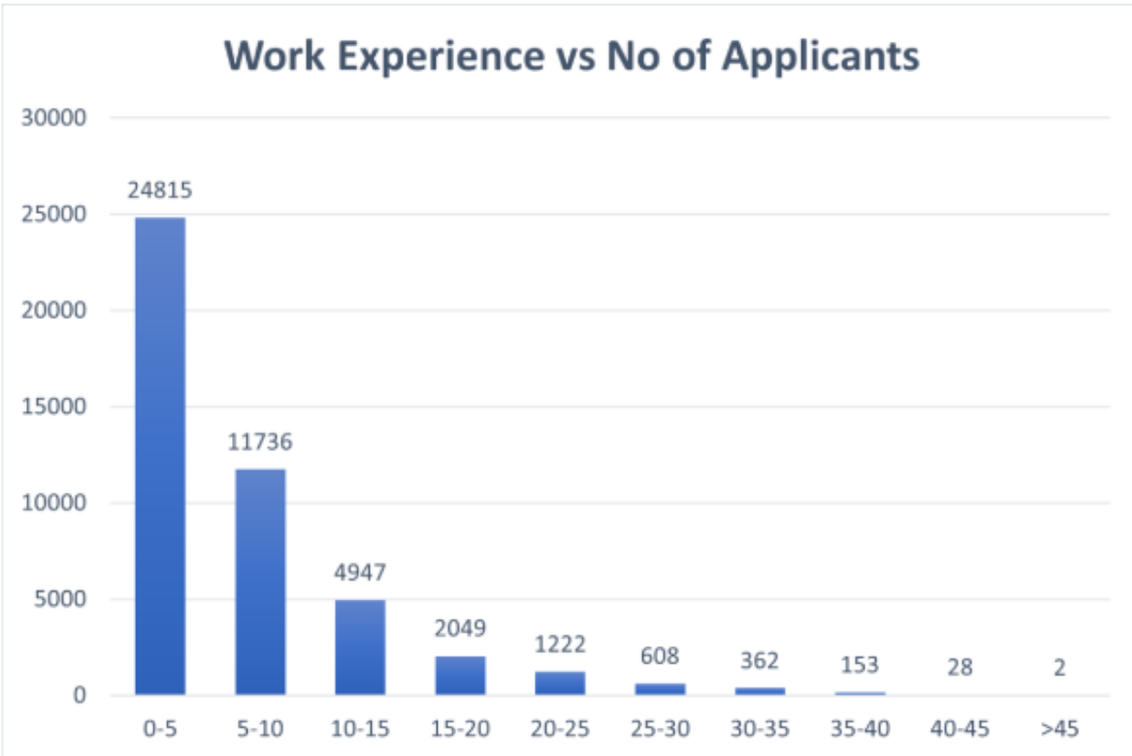


Most of the loan applicants are of age between 30-50

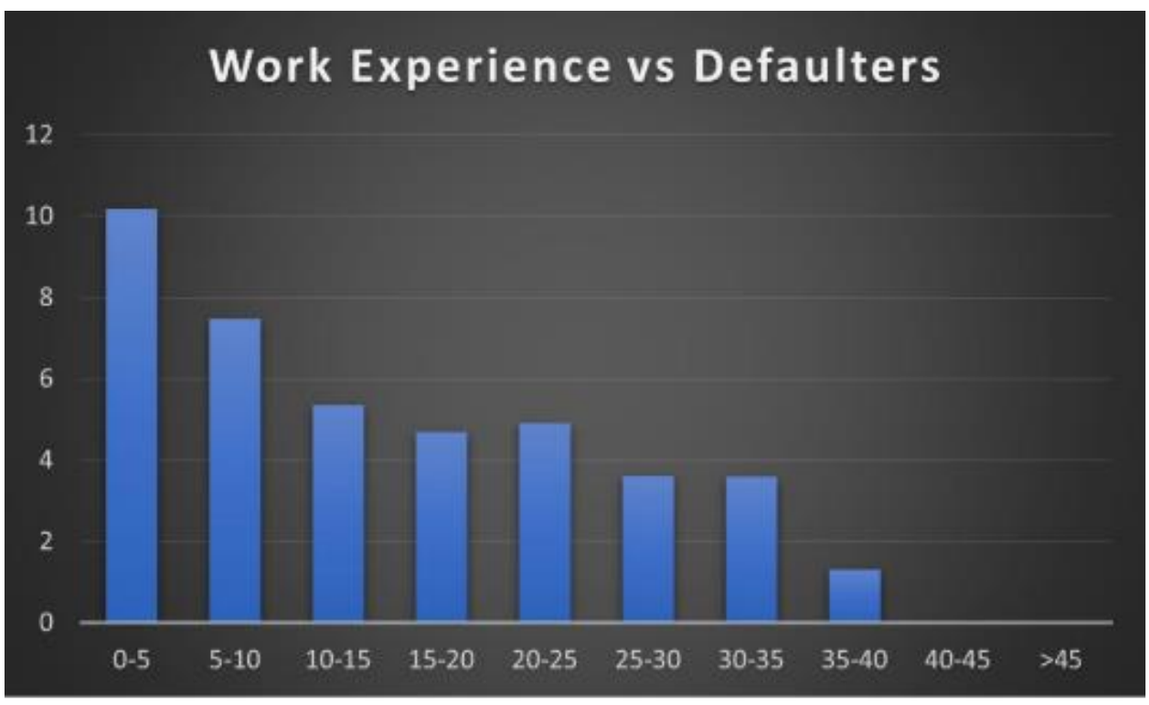


Most number of Defaulters are of the age 30-40. As per the above graph the higher the age the lower the default rate.

3.Years Employed Loan applicants and Defaulters.

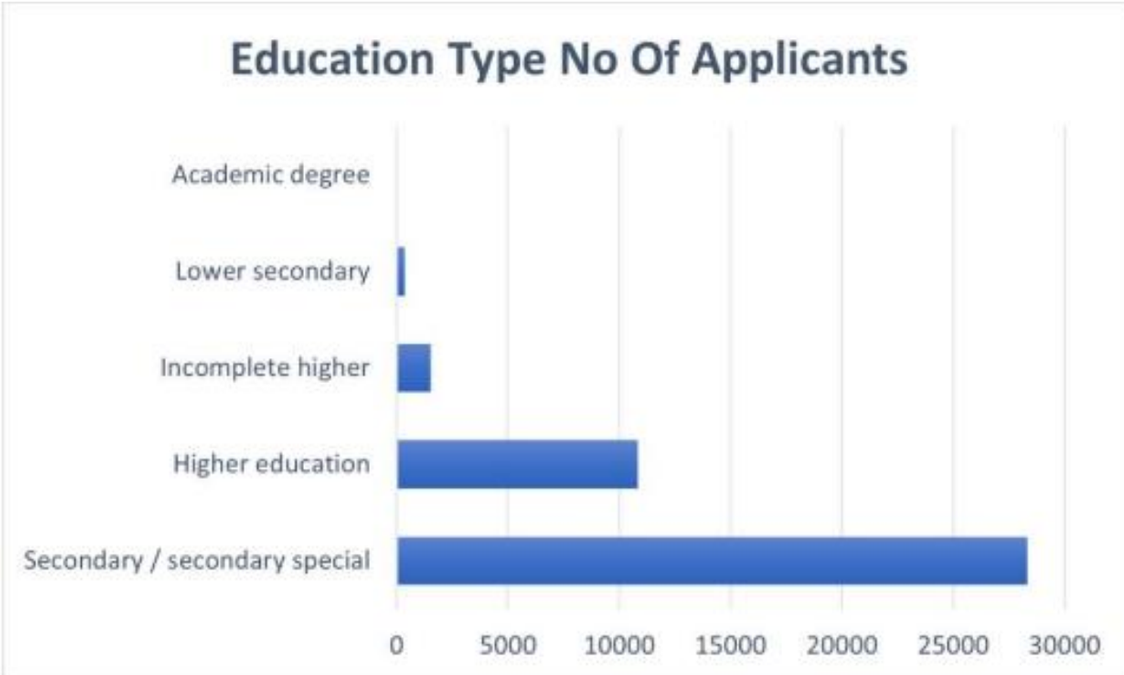


Most of the applicants are the ones who have less work Experience.

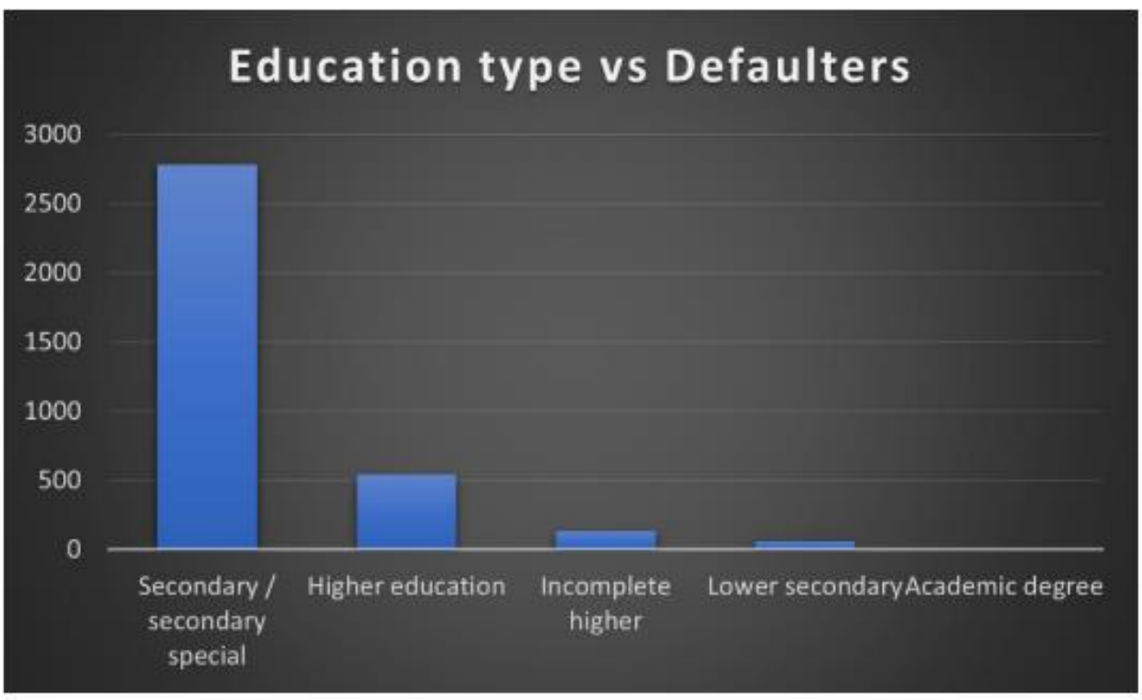


With the above graph we can say that most Experienced people are less Defaulters.

4.Education Type

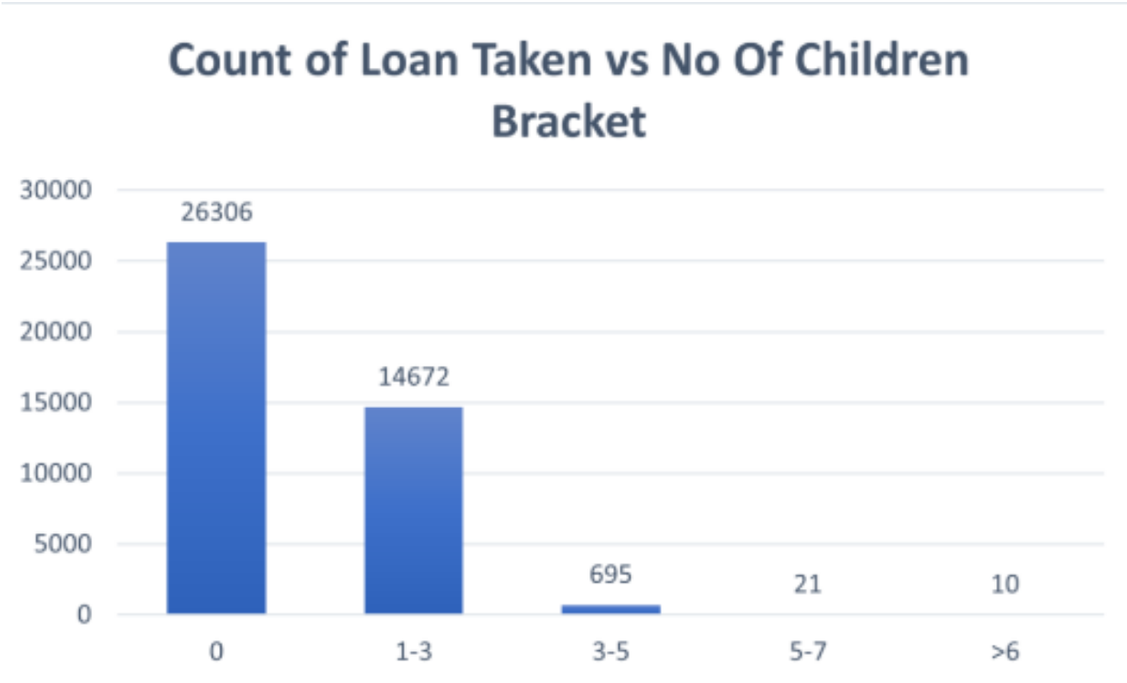


Higher Number of applicants are who completed Secondary Education

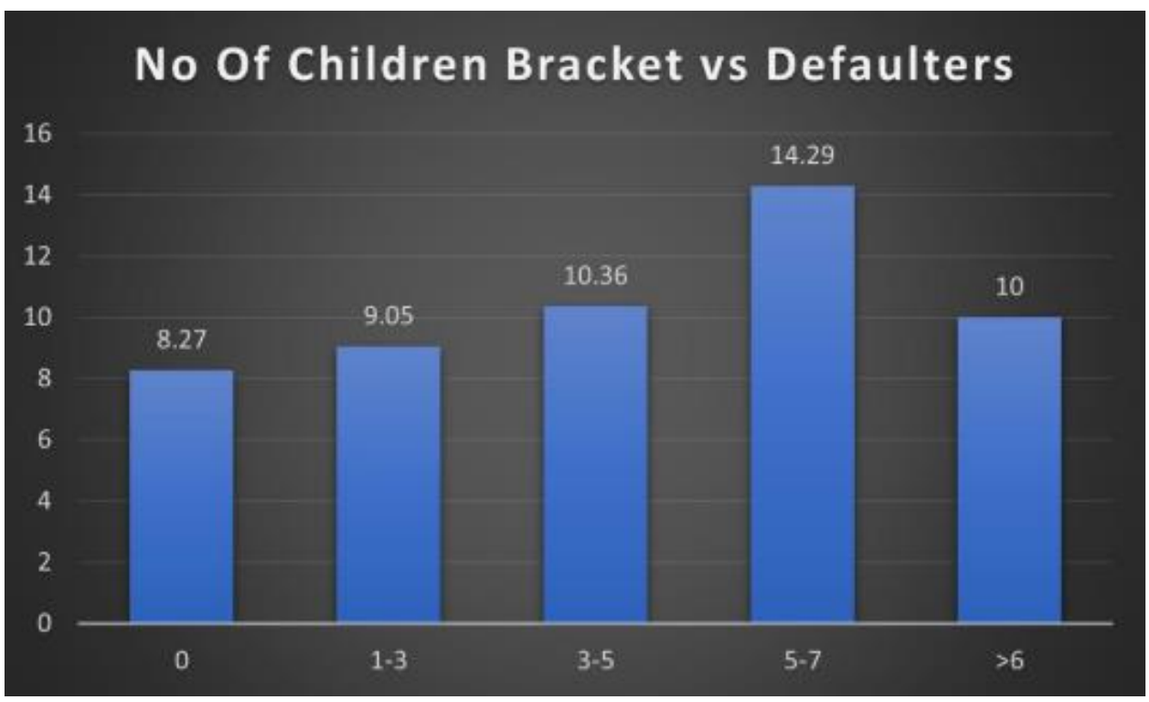


With the above graph we can see that higher the education lower the Defaulting Rate.

5. No Of Children

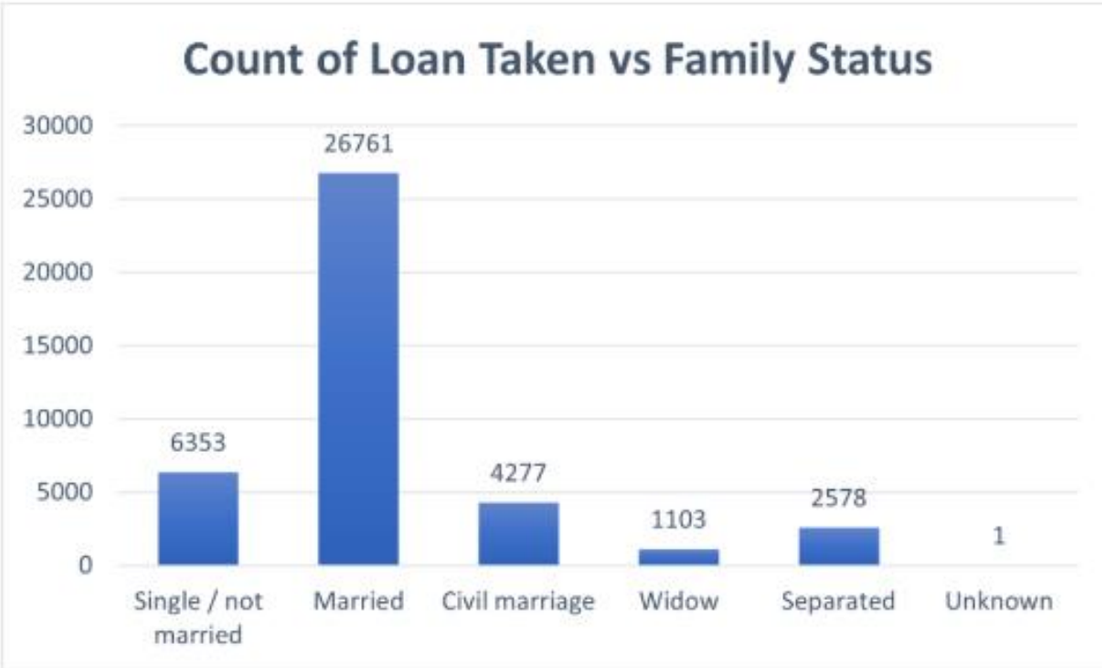


Most Number of applicants have no Children,

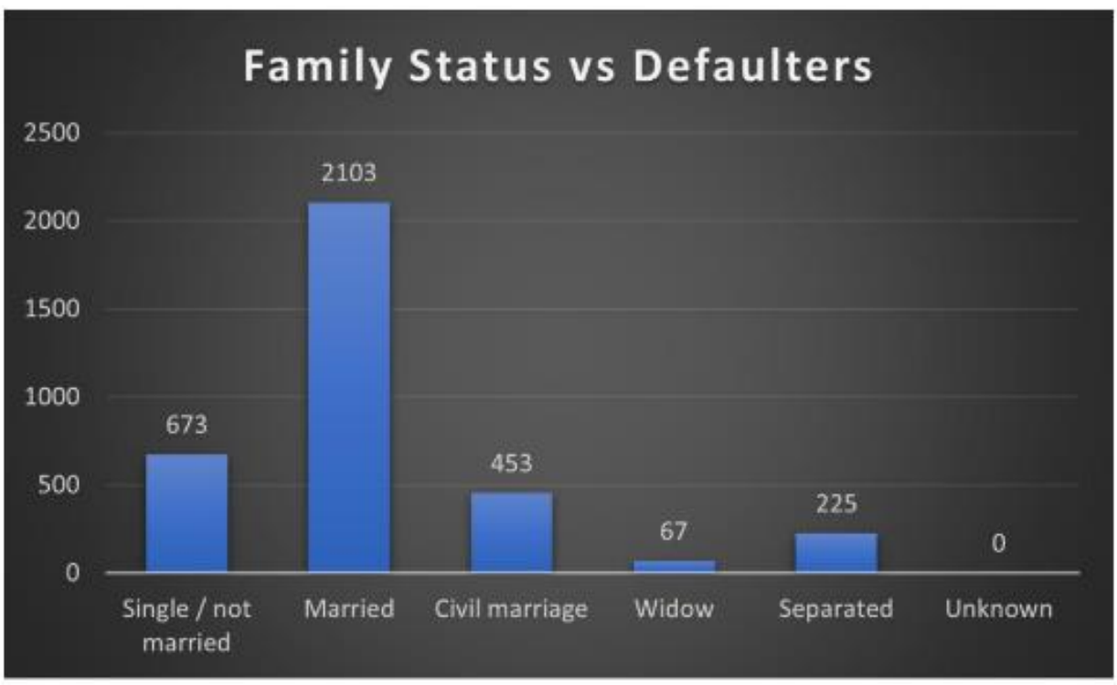


With the above graph we can see the more number of children the client has more they tend to default the loan.

5. Family Status.

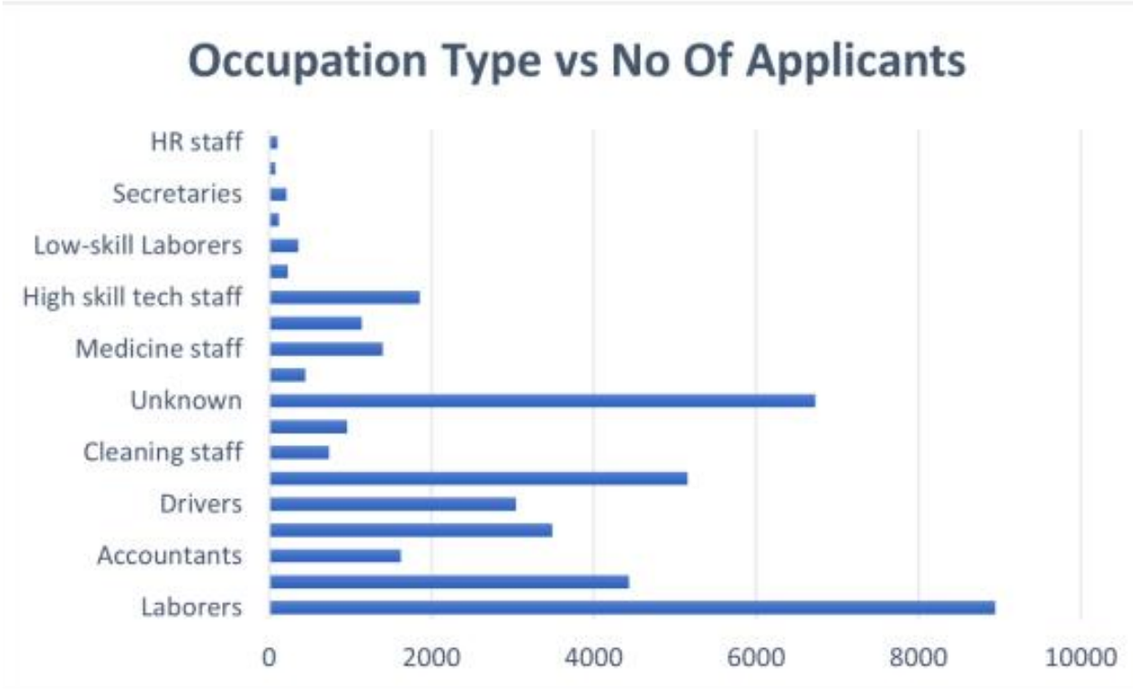


Most of the loan clients are married.

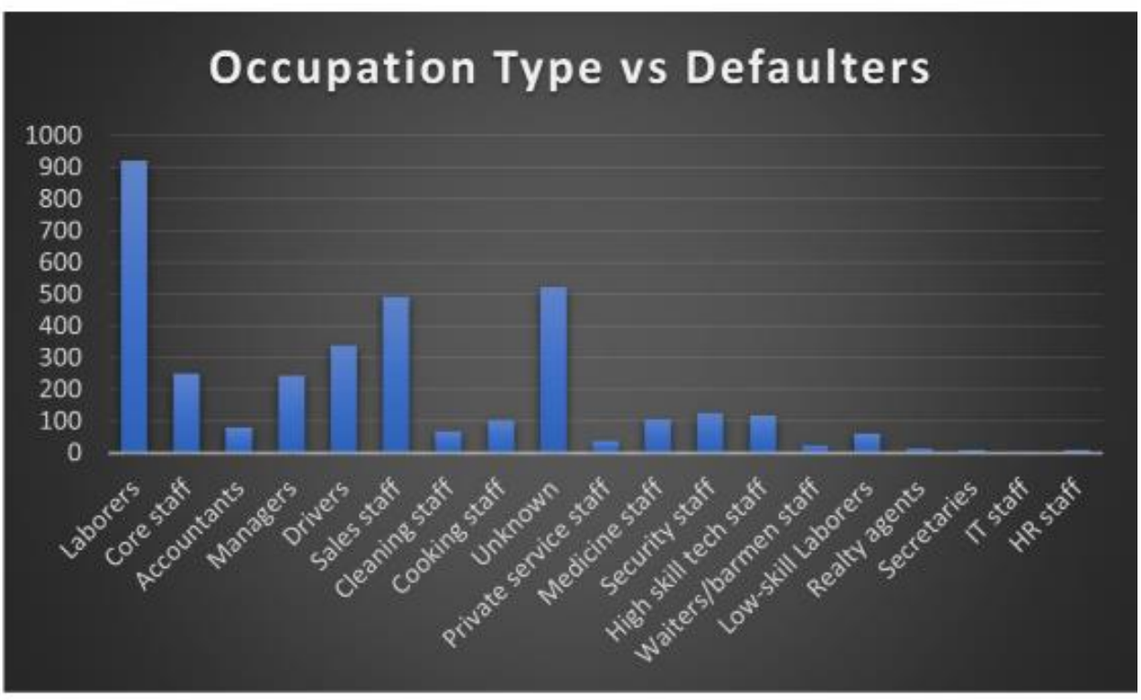


Clients who are married tend to Default the loan.

6.Ocuupation Type

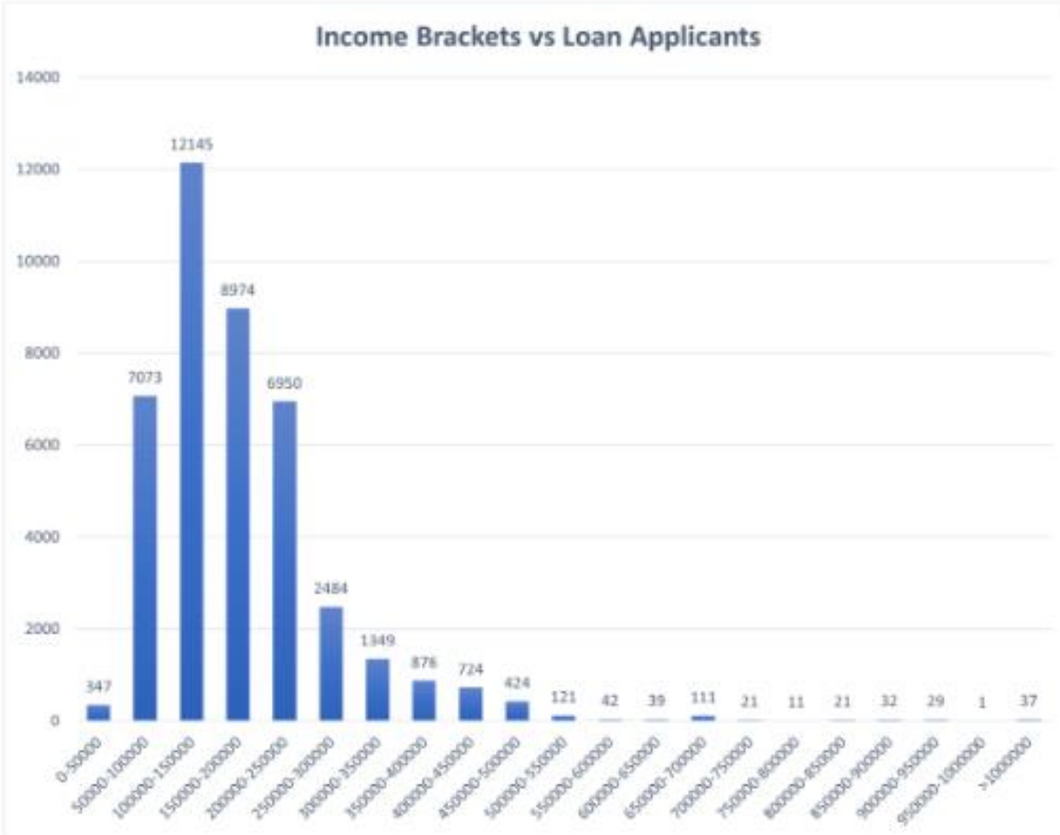


Laborers are highest no of Applicants

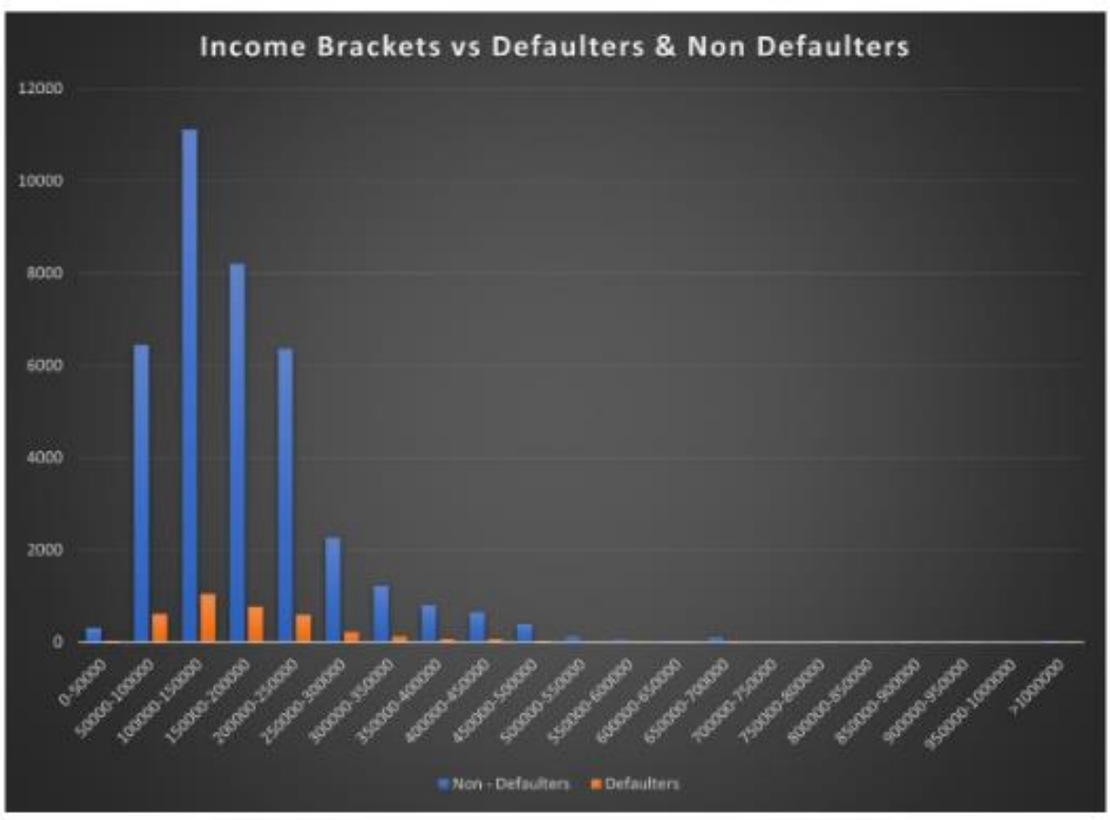


Labourers tend to Default of the loan.

7. Income Bracket

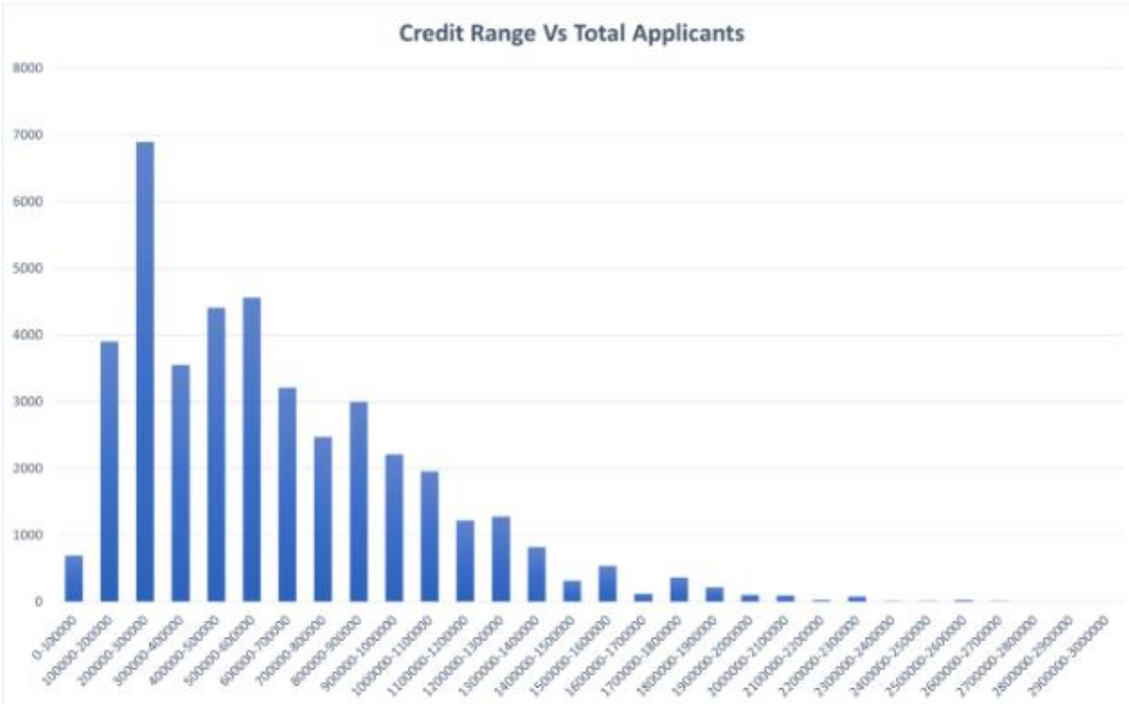


More number of applicants are from 1lakh – 1.5lakh Income range.

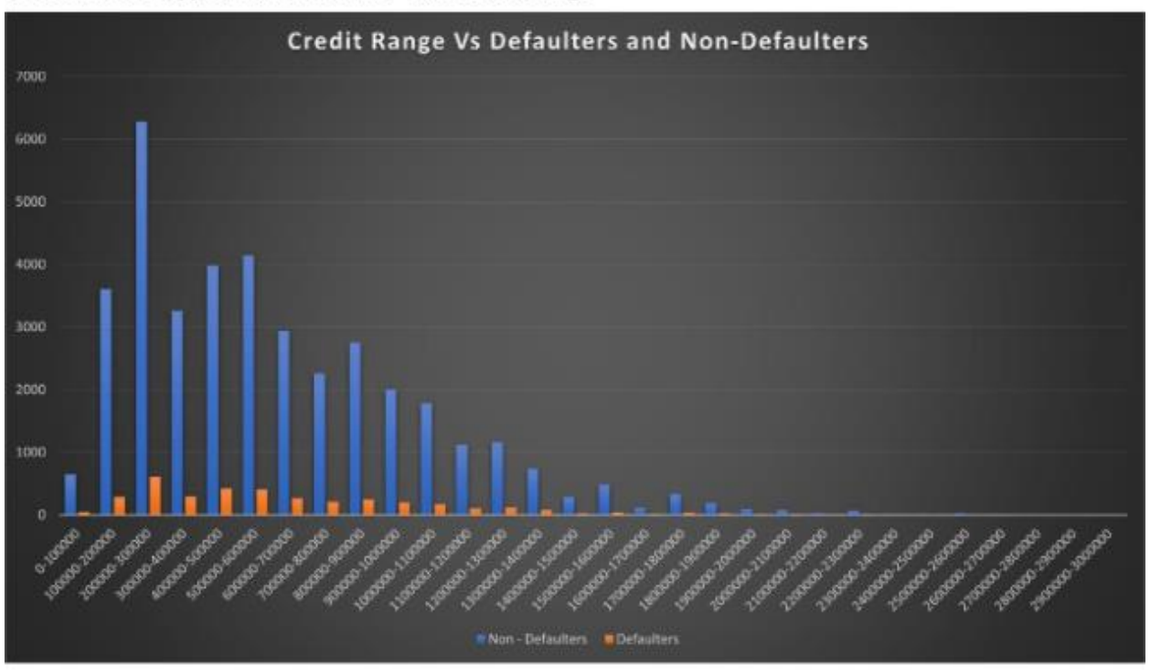


Higher number of Defaulters are form 1lakh-2lakh income Range

8. Credit Category

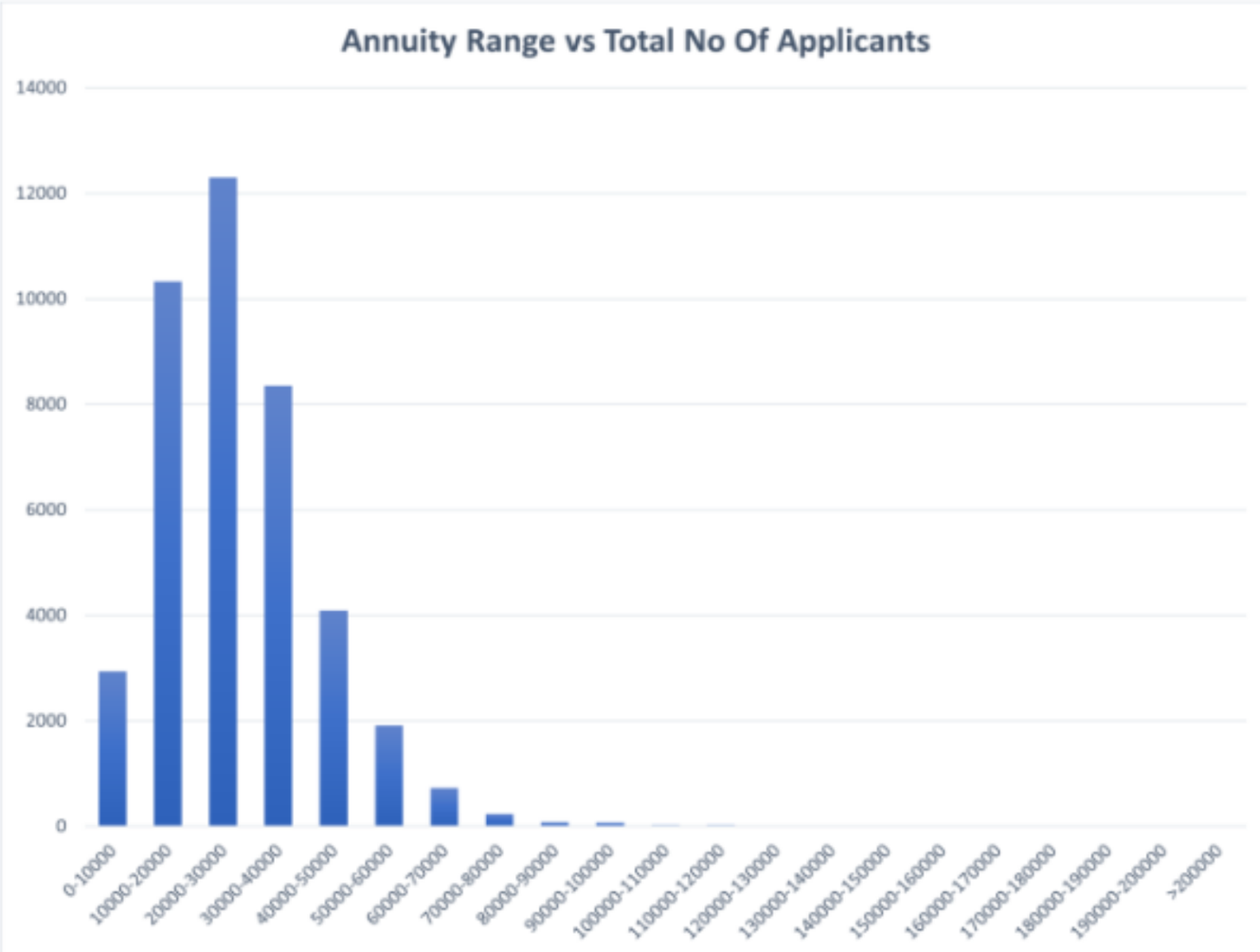


More number of applicants are of 2lakh – 3lakh credit Range

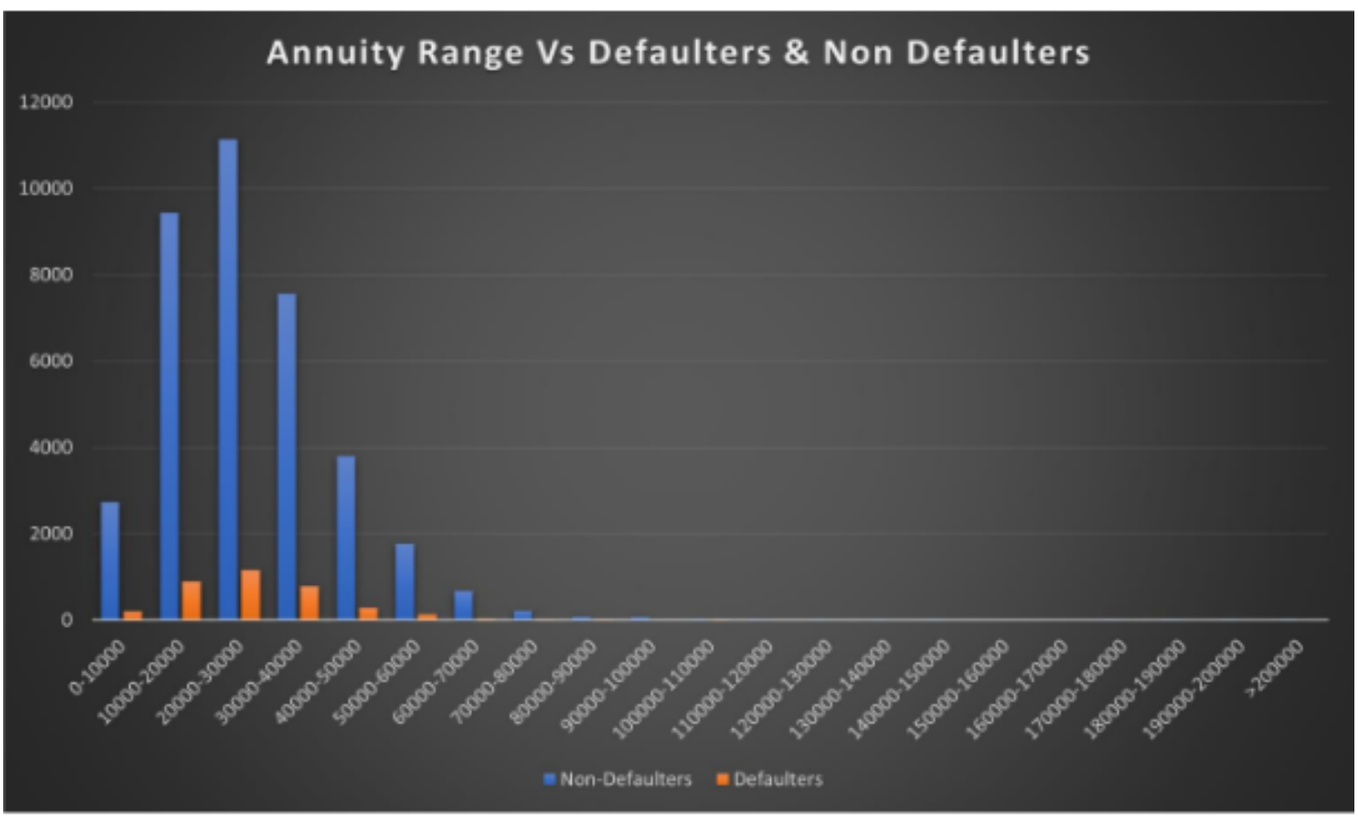


More Number of Defaulters are of 2lakh – 3lakh Credit Range.

9. Annuity Category

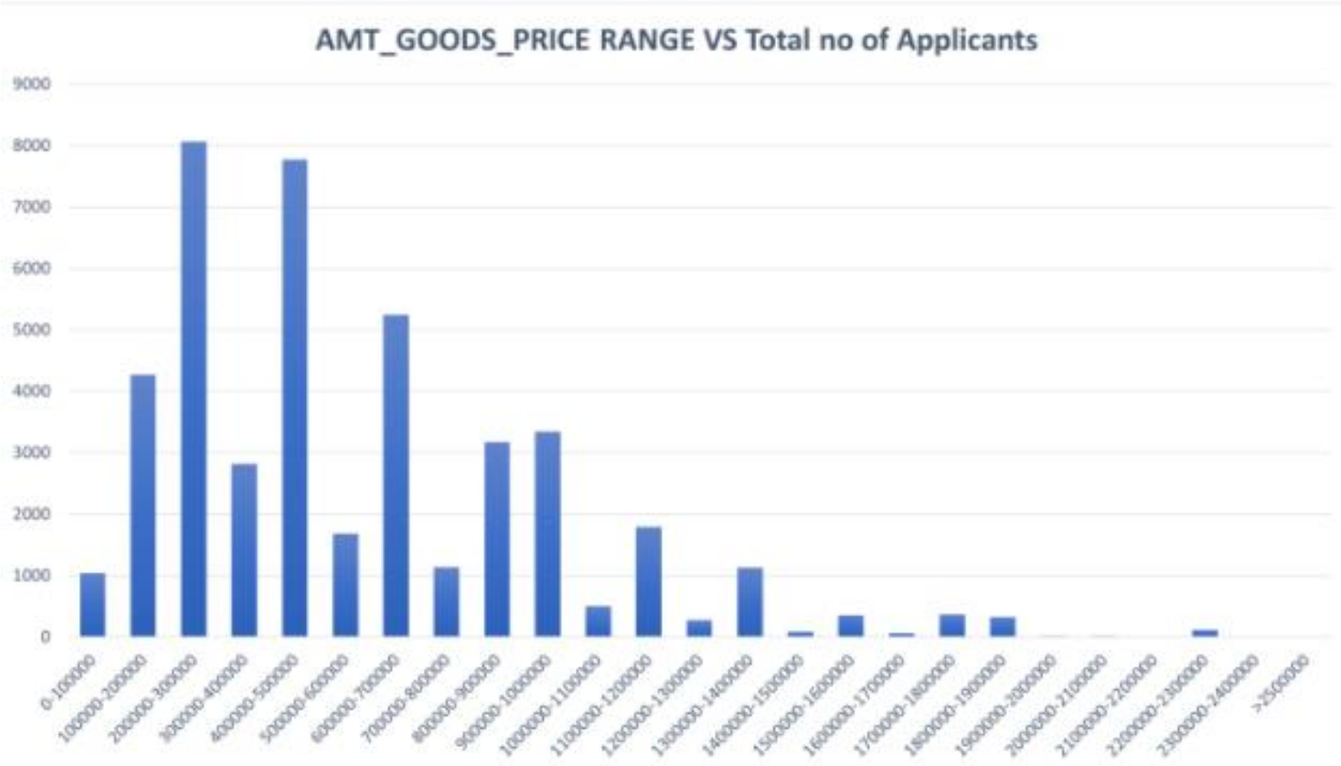


More Number of applicants are in the 20k-30k annuity range

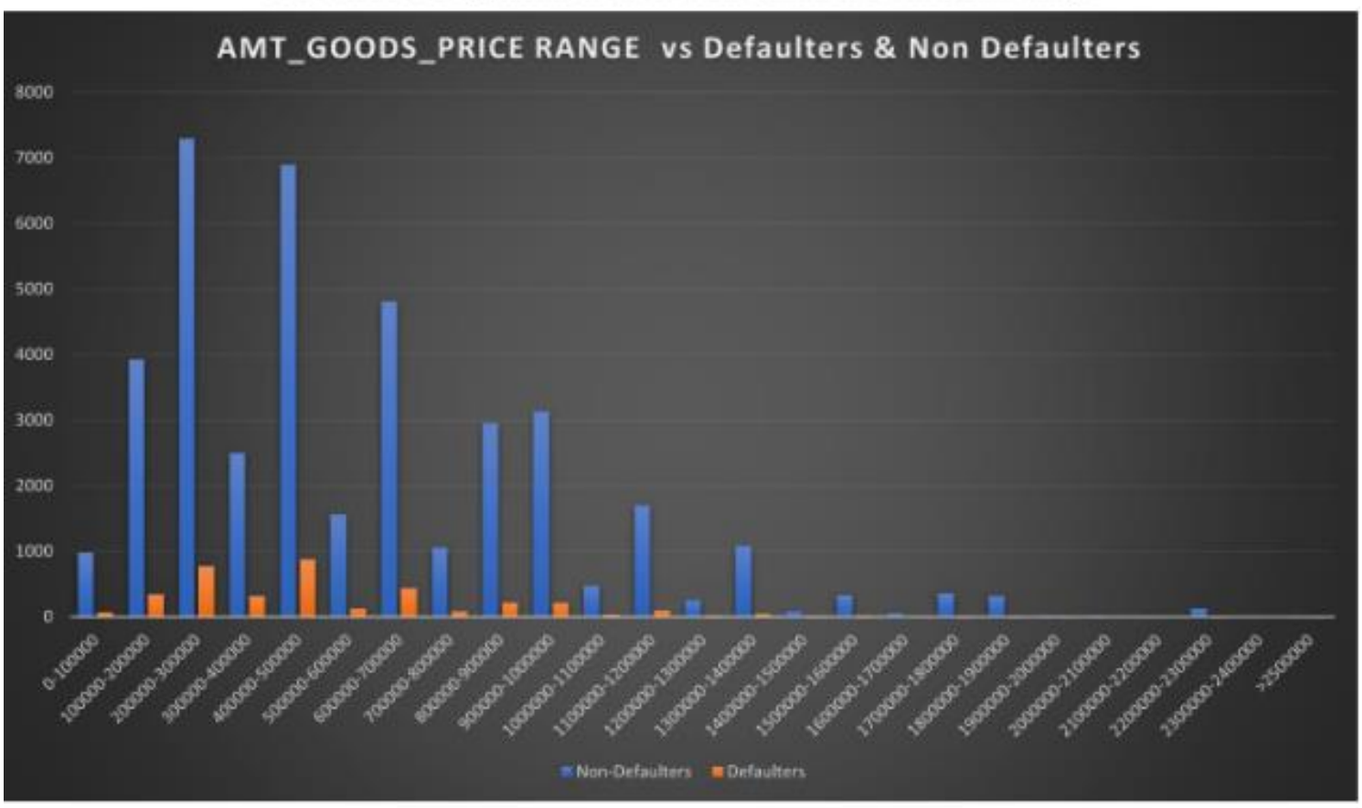


More Number of Defaulters are of 10k-30k Annuity range.

10. Goods Price Range

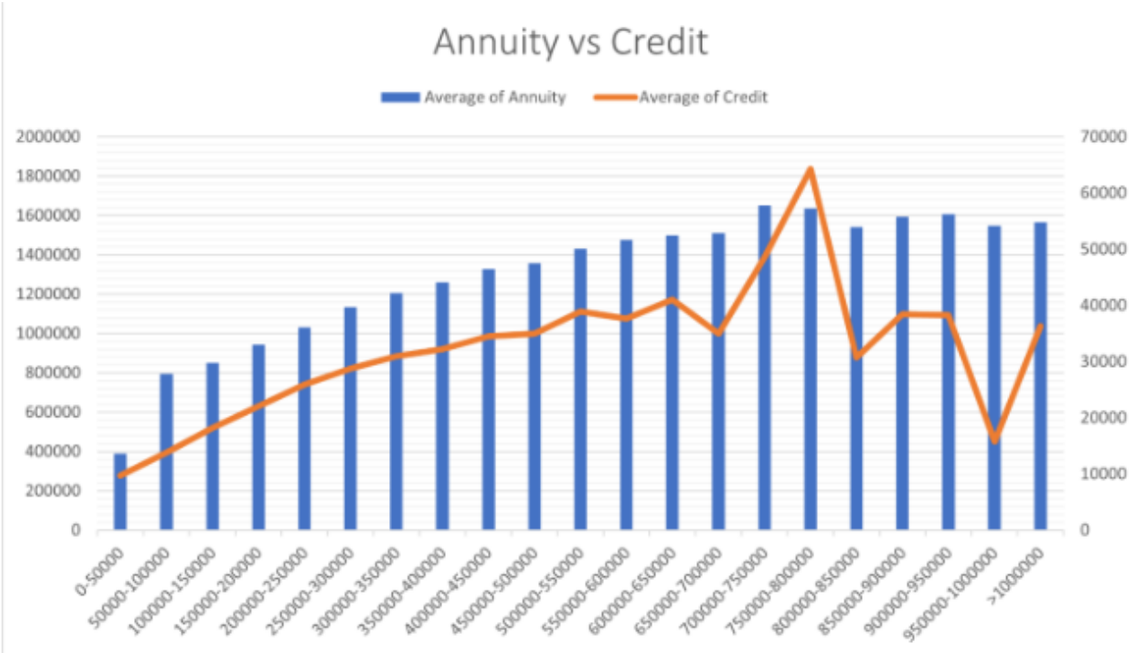


More number of applicants are from 2lakh-3lakh and 4lakh-5lakh Range



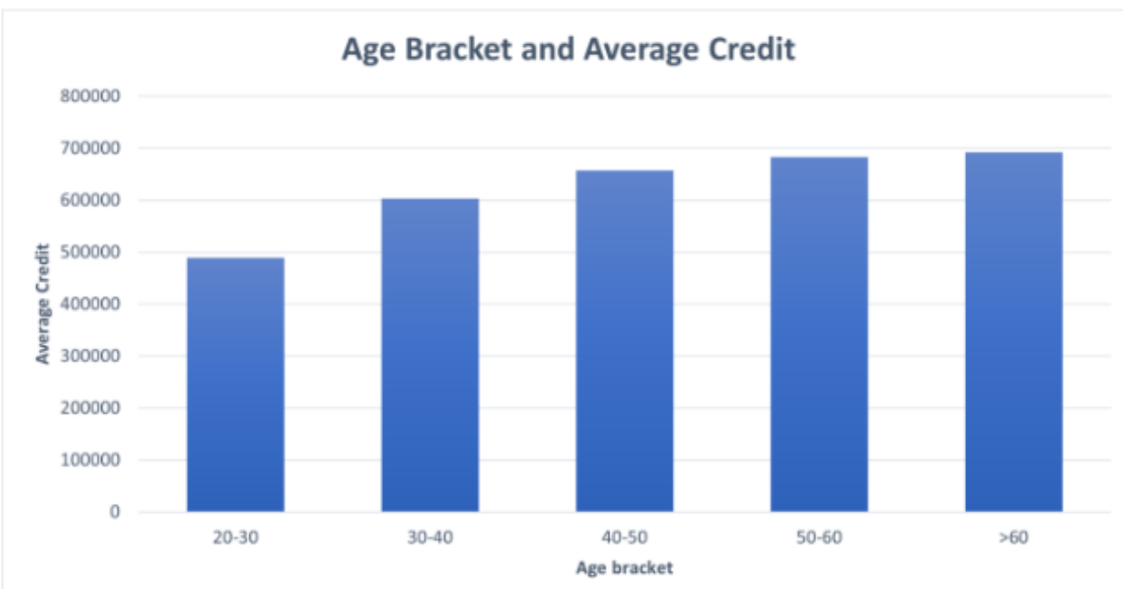
More number of Defaulters are from 2lakh-3lakh and 4lakh-5lakh Range

11. Annuity vs Credit and Income Range Bi – variate.



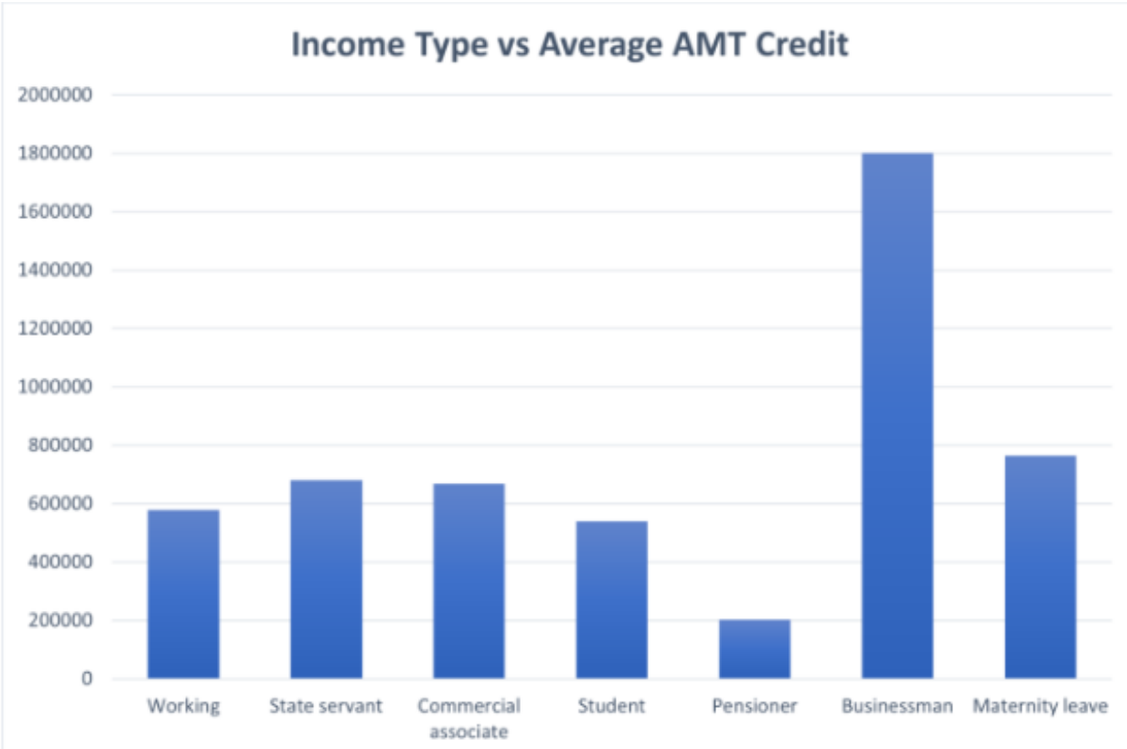
As the income of the client increases their respective annuity and credit increases and caps out at 7.5Lakh-8lakhs,

12.Age Bracket vs Average Credit



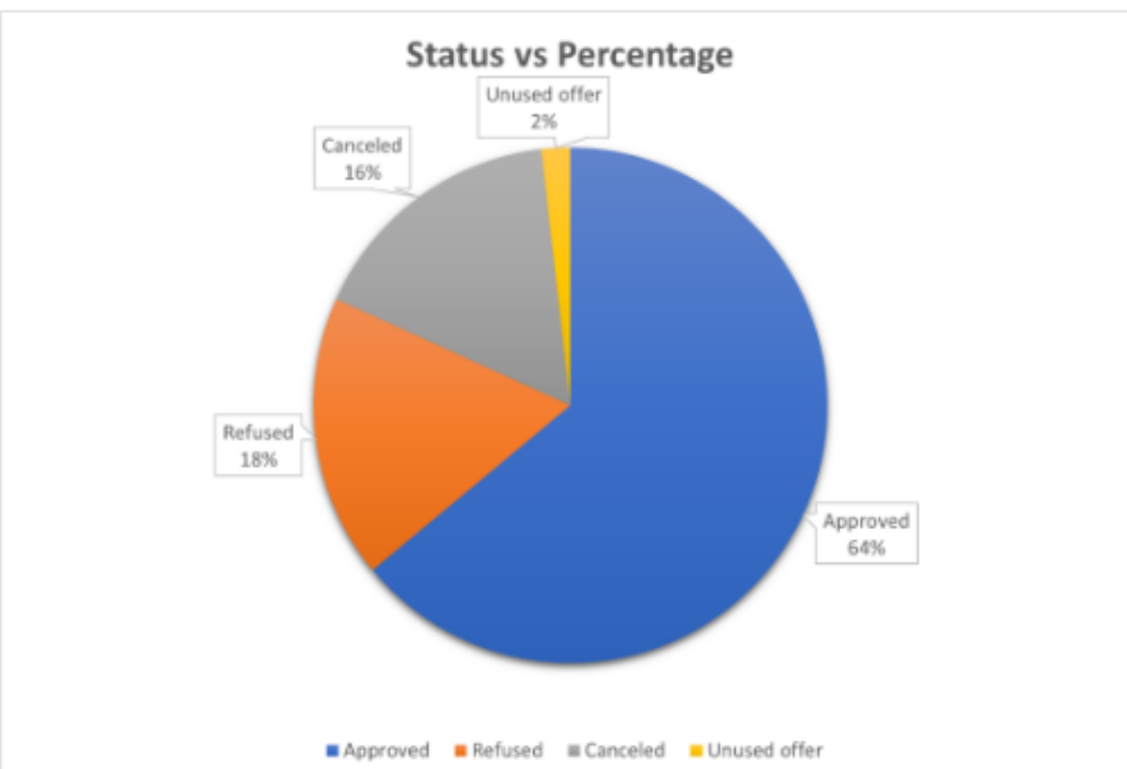
Credit Amount increases as the client gets older and older.

13.Income type vs Credit

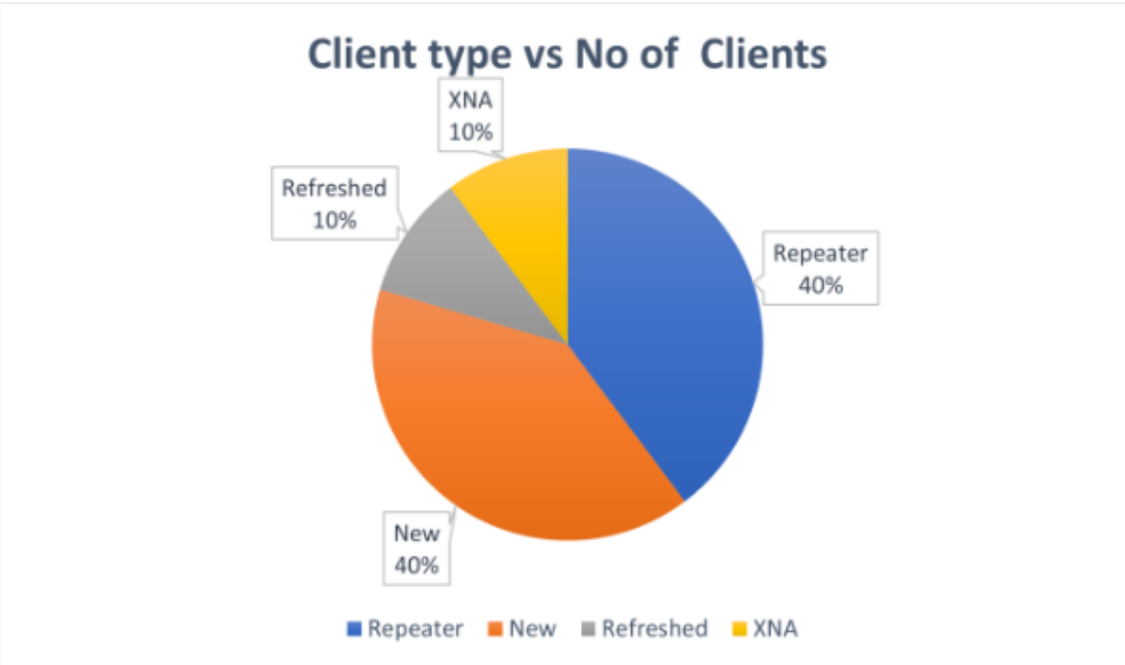


Clients who are businessman has high credit amount.

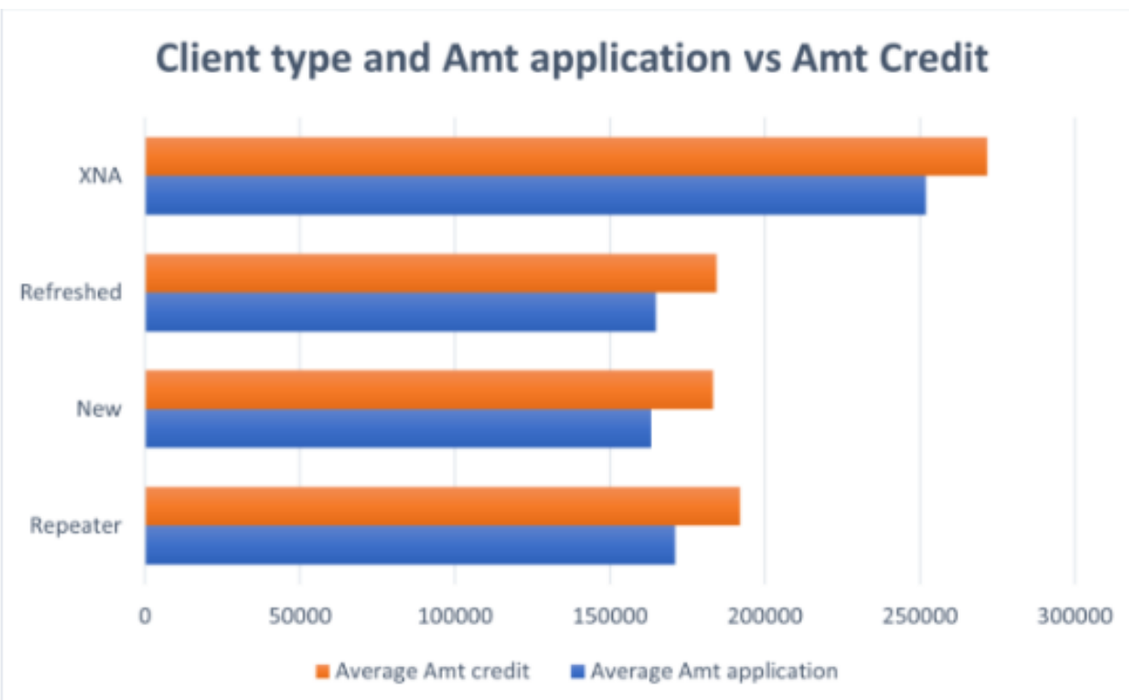
14. Contract type – Previous Application



15. Client Type - Previous Application



16.Client type and application type vs Amt Credit.



E. Identify Top Correlations for Different Scenarios: Understanding the correlation between variables and the target variable can provide insights into strong indicators of loan default.

	Top Correlation for Non Defaulters		
Rank	Variable 1	Variable 2	Correlation
1	OBS_60_CNT_SOCIAL_CIRCLE	OBS_30_CNT_SOCIAL_CIRCLE	0.99834471
2	AMT_GOODS_PRICE	AMT_CREDIT	0.986687101
3	REGION_RATING_CLIENT_W_CITY	REGION_RATING_CLIENT	0.950291816
4	CNT_FAM_MEMBERS	CNT_CHILDREN	0.893726572
5	LIVE_REGION_NOT_WORK_REGION	REG_REGION_NOT_WORK_REGION	0.859968146
6	DEF_60_CNT_SOCIAL_CIRCLE	DEF_30_CNT_SOCIAL_CIRCLE	0.853001934
7	LIVE_CITY_NOT_WORK_CITY	REG_CITY_NOT_WORK_CITY	0.815622188
8	AMT_GOODS_PRICE	AMT_ANNUITY	0.76701179
9	AMT_ANNUITY	AMT_CREDIT	0.762033985

	Top Correlation of Defaulters		
Rank	Variable 1	Variable 2	Correlation
1	DEF_60_CNT_SOCIAL_CIRCLE	OBS_30_CNT_SOCIAL_CIRCLE	0.99803952
2	AMT_GOODS_PRICE	AMT_CREDIT	0.982100674
3	REGION_RATING_CLIENT_W_CITY	REGION_RATING_CLIENT	0.948022037
4	CNT_FAM_MEMBERS	CNT_CHILDREN	0.897062042
5	OBS_60_CNT_SOCIAL_CIRCLE	DEF_30_CNT_SOCIAL_CIRCLE	0.891518651
6	LIVE_REGION_NOT_WORK_REGION	REG_REGION_NOT_WORK_REGION	0.805585864
7	LIVE_CITY_NOT_WORK_CITY	REG_CITY_NOT_WORK_CITY	0.773132306
8	AMT_GOODS_PRICE	AMT_ANNUITY	0.746639466
9	AMT_ANNUITY	AMT_CREDIT	0.745137729

Insights:

• With the analysis, we found out few key factors which helps us analyze if the client will default or re-pay the loan.

• As the age and years of Experience Increases, the chances of default decreases. We can prioritize older and Experienced clients

• Clients with high level of education tends to default the loan less.

• Female Clients are higher than male clients applying for loan. Female clients default more than male.

• As age Increases the amount taken by client are higher and they tend to default on the loan less

Results:

• In this case study, I applied the EDA to analyze patterns in the data.

• Before starting the project, I conduct brief research on risk analytics in banking and financial services.

• I comprehend how data is employed to reduce the risk of financial losses when extending loans to customers.

• This case study has assisted me in acquiring the skill of summarizing extensive datasets to extract valuable insights.

• I gained knowledge about data imbalance, outliers, and the underlying factors influencing the datasets.

Analysing the Impact of Car Features on Price and Profitability

Project Description

The automotive industry is shifting rapidly with a focus on fuel efficiency, environmental sustainability, and technological innovation. Electric and hybrid vehicles are becoming more popular, while traditional gasoline-powered cars remain dominant. As a Data Analyst, we can help car manufacturers optimize pricing and product development decisions by analysing the relationship between a car's features, market category, and pricing. By using data analysis techniques such as regression analysis and market segmentation, the manufacturer could develop a pricing strategy that balances

consumer demand with profitability and identify which product features to focus on in future product development efforts

Tech Stack Used

Microsoft Excel and Microsoft Power Point

Understanding Data

Our data set contains details of Cars and their features and MSRP of Cars.

Dataset is in the Format of csv.

Dataset Contains 11915 rows and 16 Columns

Data Cleaning

Our Dataset Contained Blank Values in

Fuel type - Imputed with online and mode data

Fuel Type			Imputation values
Make	Model	Year	Fuel Type
Suzuki	Verona	2004	Regular Unleaded
Suzuki	Verona	2004	Regular Unleaded
Suzuki	Verona	2004	Regular Unleaded

No of Doors – Imputed with Values found for exact model or mode for particular car

No Of Doors		
Make	Model	No of Doors
Ferrari	FF	2
Tesla	Model S	4
Tesla	Model S	4
Tesla	Model S	4
Tesla	Model S	4
Tesla	Model S	4

Engine HP – Had 69 Empty Values, Imputed with mode and Values found on the Internet

•Engine Cylinders – EV's Don't have a cylinder coz they run using motors. RX7 & RX8 Use a different type of engine called a rotary engine.

Had an outlier in highway mpg – 354, Imputed with 34 which is mode of particular model.

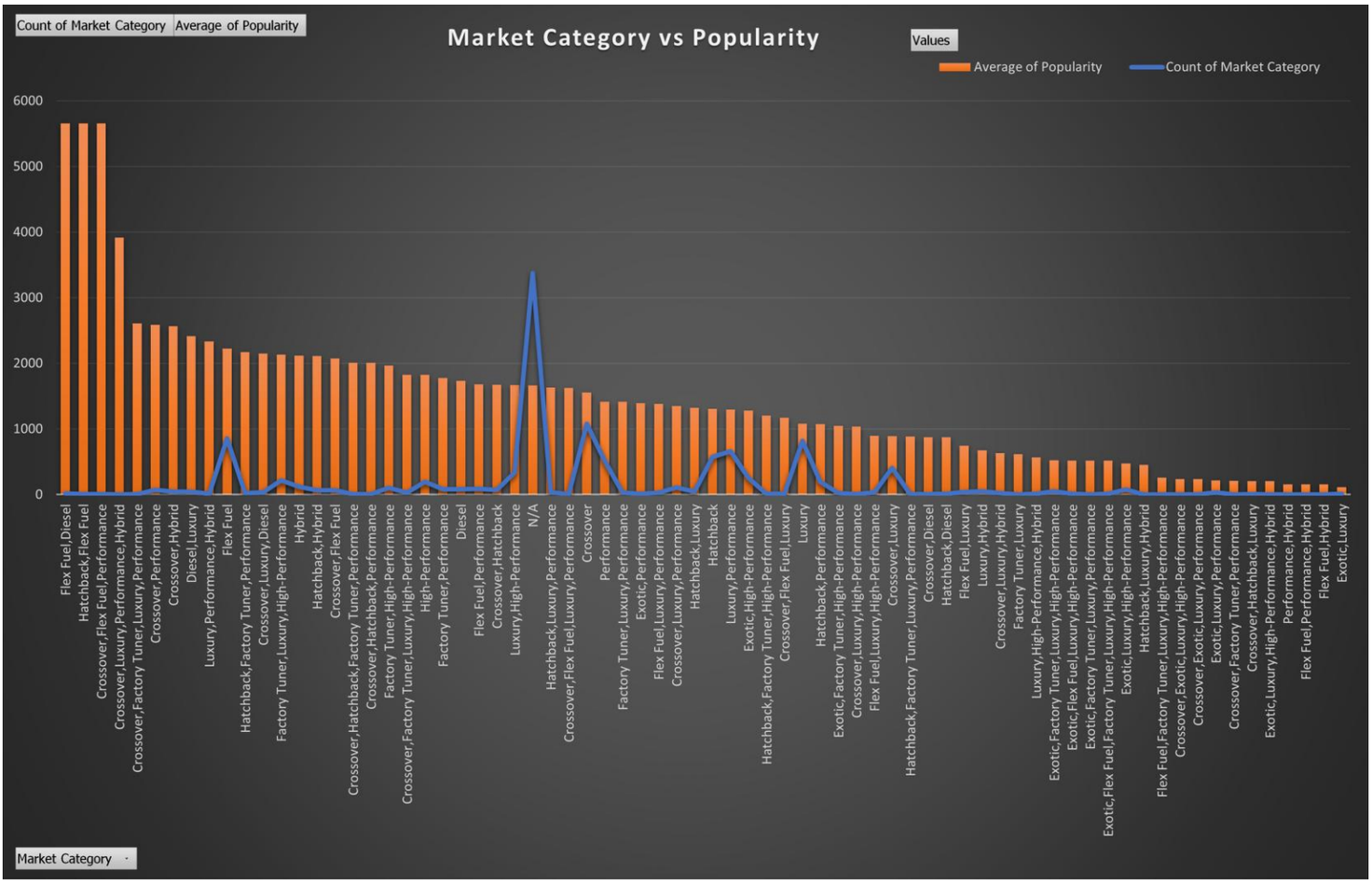
Used a box plot to find the outlier

Engine Fuel Type – Transformed Engine Fuel type.

Removed 715 duplicate entries

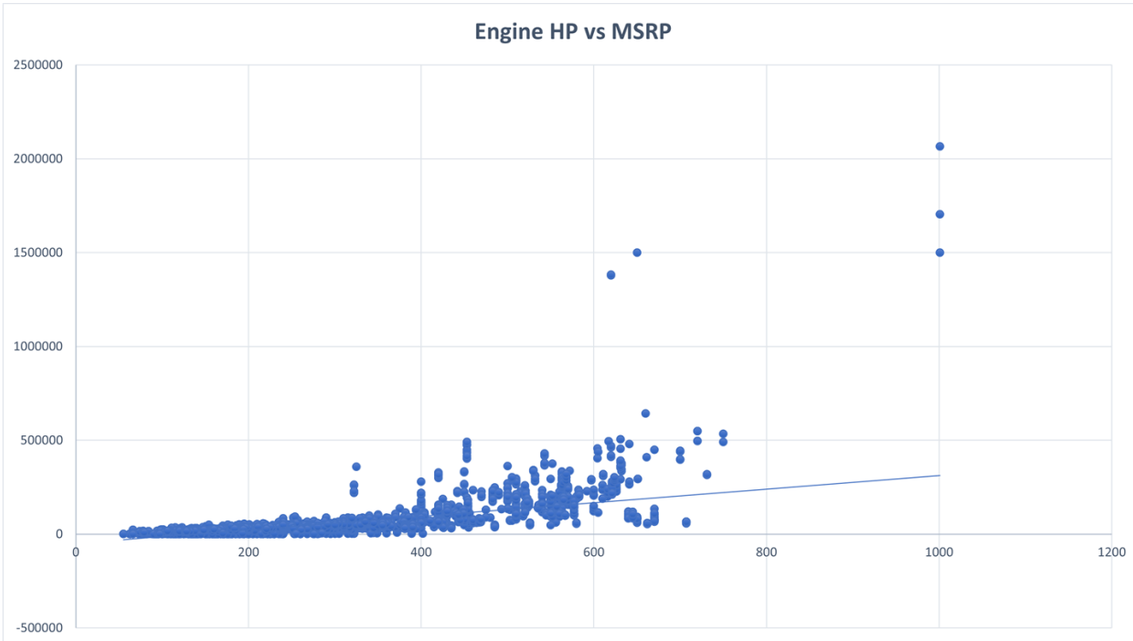
Data Analysis

Task 1: How does the popularity of a car model vary across different market categories?



Crossover's are the highest selling cars but Flex Fuel, Diesel have the highest Popularity Score and Exotic cars the least.

Task 2: What is the relationship between a car's engine power and its price?



Engine Power and MSRP have a positive linear relation. The Higher the Engine power,

The higher the price of the car.

Exception: tesla has electric motors with which the car can make close 1000hp and also cost less

Task 3 :Which car features are most important in determining a car's price?

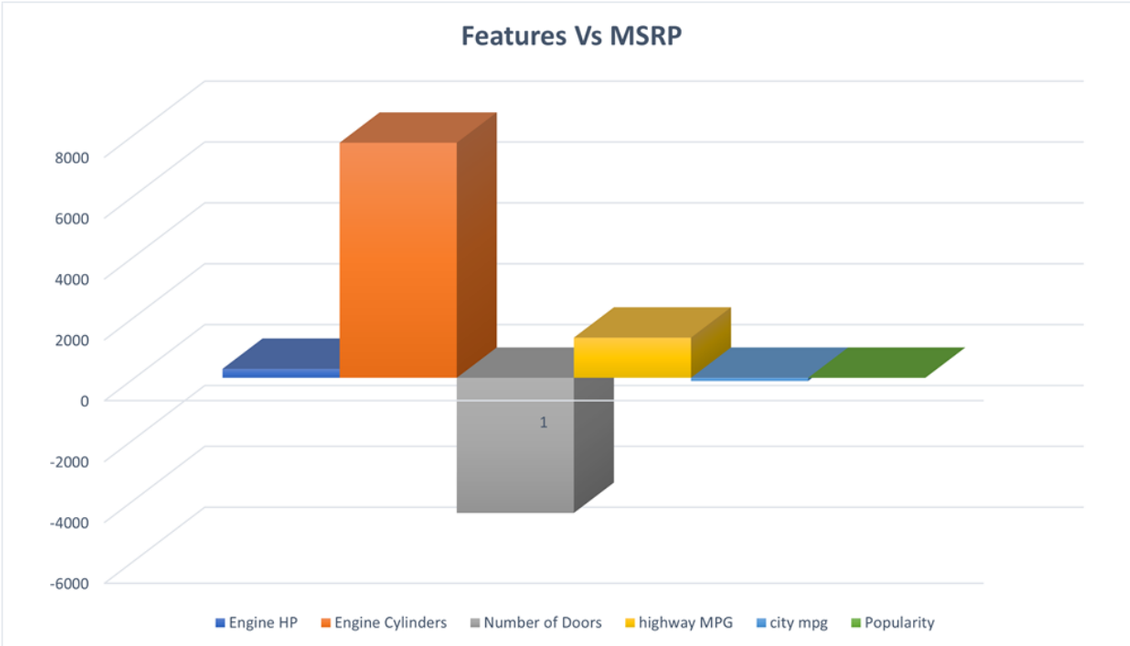
SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.678764735
R Square	0.460721565
Adjusted R Square	0.46044982
Standard Error	44152.58877
Observations	11914

ANOVA

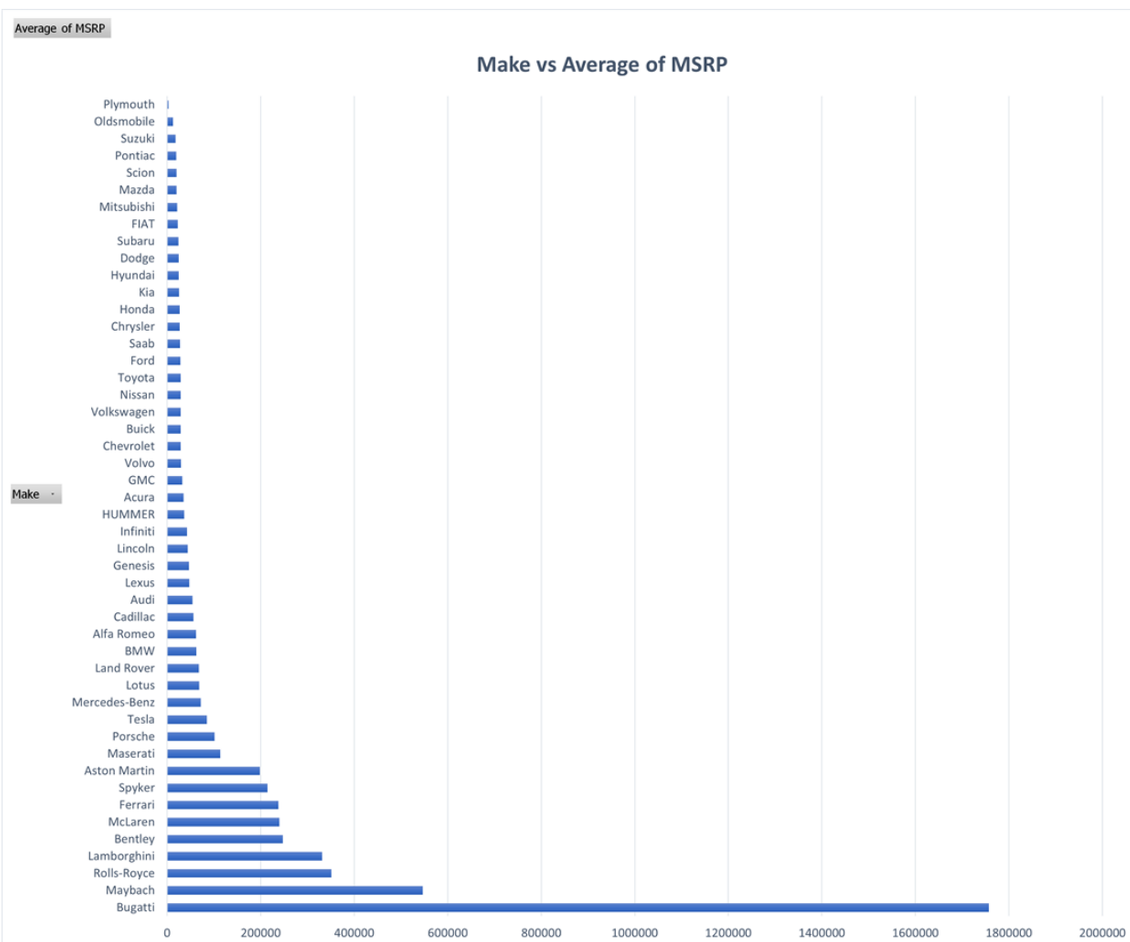
	df	*SS*	*MS*	*F*	*Significance F*
Regression	6	1.98308E+13	3.30513E+12	1695.417223	0
Residual	11907	2.32121E+13	1949451095		
Total	11913	4.30429E+13			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	-89298.13533	3870.286734	-23.07274408	2.874E-115	-96884.52911	-81711.74155	-96884.52911	-81711.74155
Engine HP	294.7961646	5.686434758	51.84200245	0	283.6498243	305.942505	283.6498243	305.942505
Engine Cylinders	7724.137619	433.9108478	17.80120884	5.56225E-70	6873.601526	8574.673711	6873.601526	8574.673711
Number of Doors	-4440.912424	464.7194298	-9.556115236	1.46079E-21	-5351.838366	-3529.986482	-5351.838366	-3529.986482
highway MPG	1319.31517	156.6258871	8.423353218	4.0768E-17	1012.302864	1626.327476	1012.302864	1626.327476
city mpg	-103.2446228	134.3180693	-0.768657734	0.442111767	-366.5299643	160.0407188	-366.5299643	160.0407188
Popularity	-3.162007423	0.281613091	-11.22819758	4.15114E-29	-3.714015051	-2.609999796	-3.714015051	-2.609999796



Engine Cylinder and Highway MPG have the highest positive coefficient in determining MSRP.

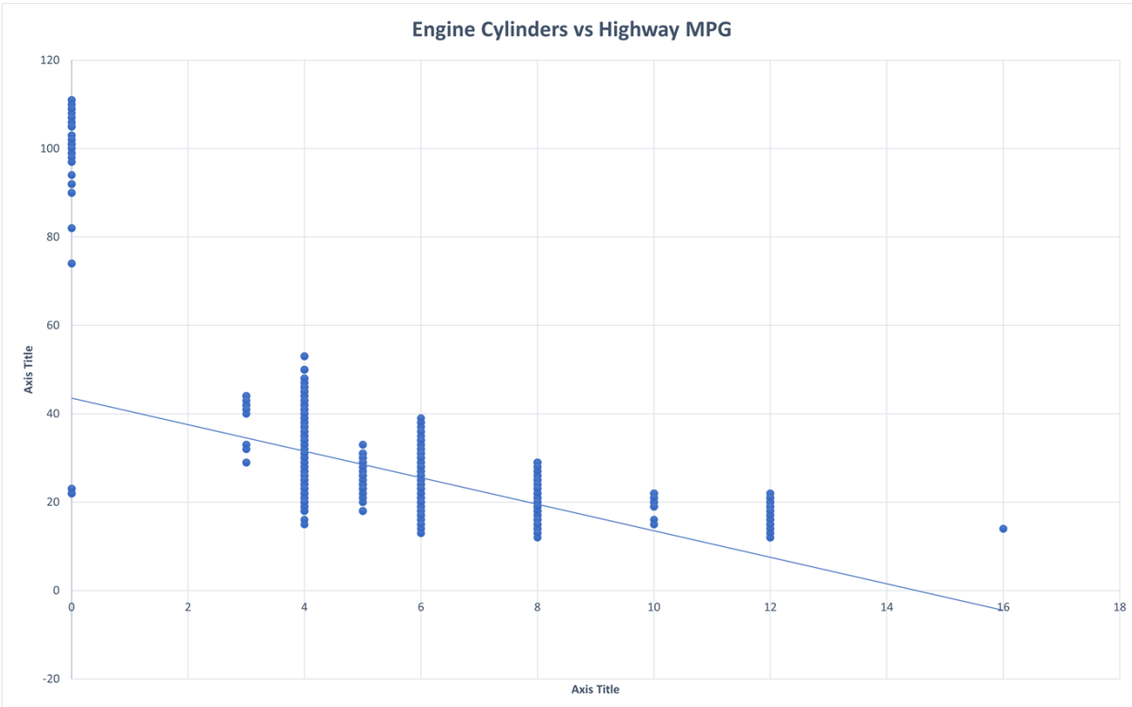
Task 4: How does the average price of a car vary across different manufacturers?



Luxury/Sports cars are some of the most expensive cars

Everyday/Cars for the masses are the once which are cheaper.

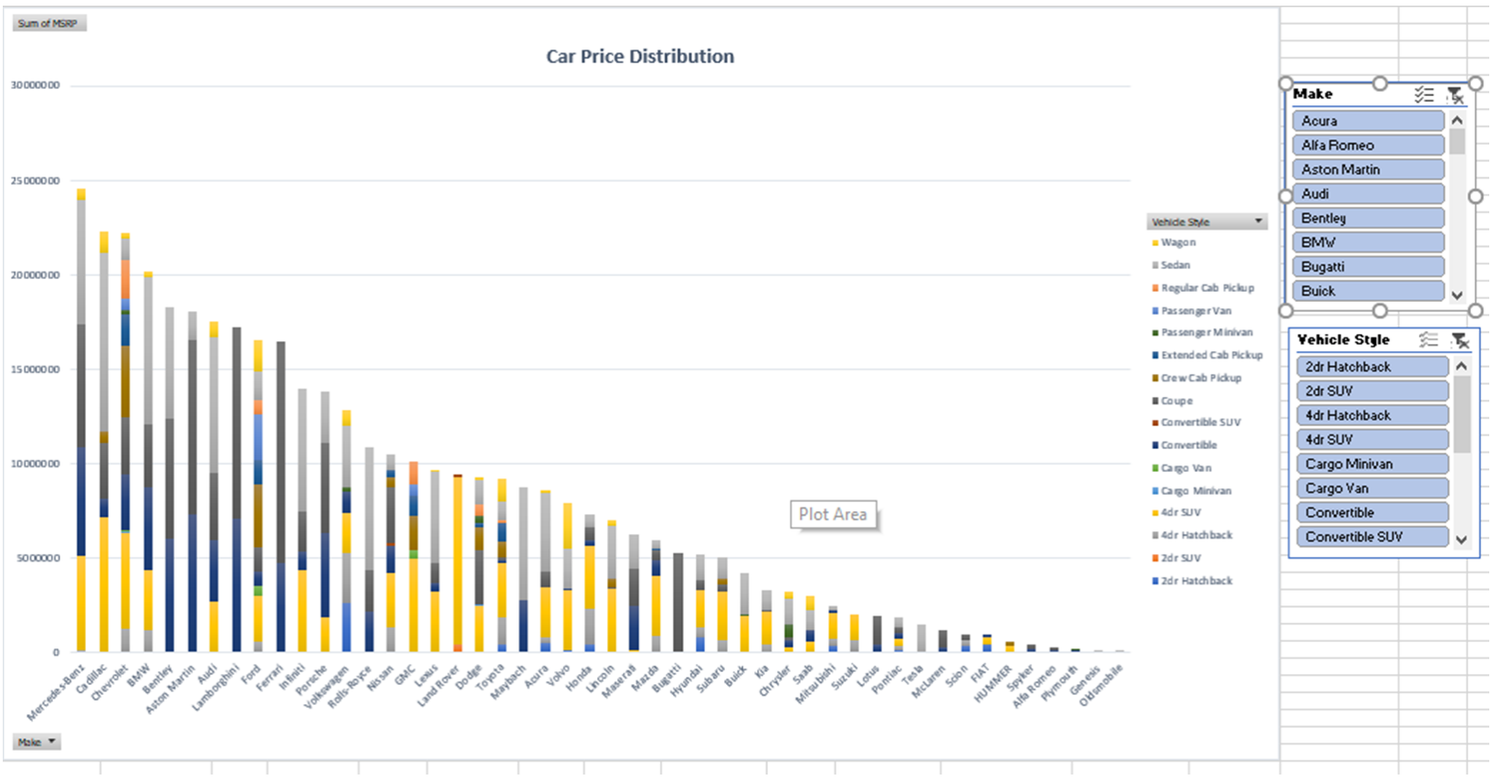
Task 5 : What is the relationship between fuel efficiency and the number of cylinders in a car's engine?



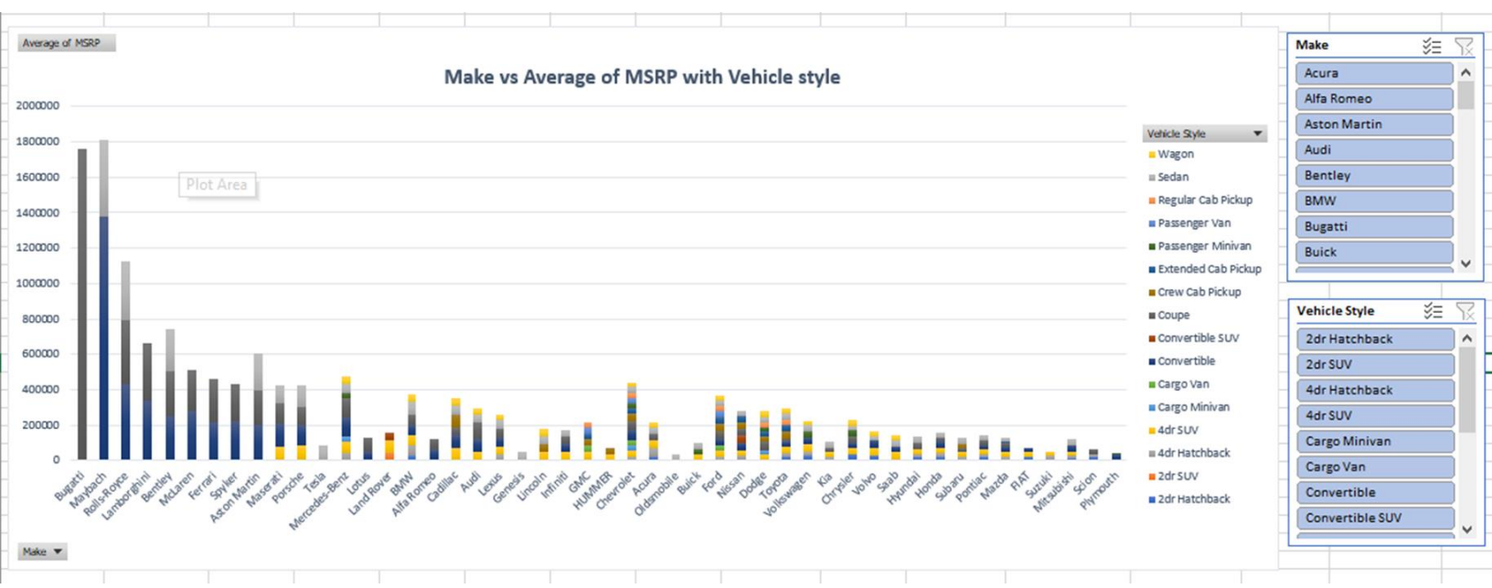
We clearly see that as the Number of cylinders increase, the efficiency decreases

Building the Dashboard

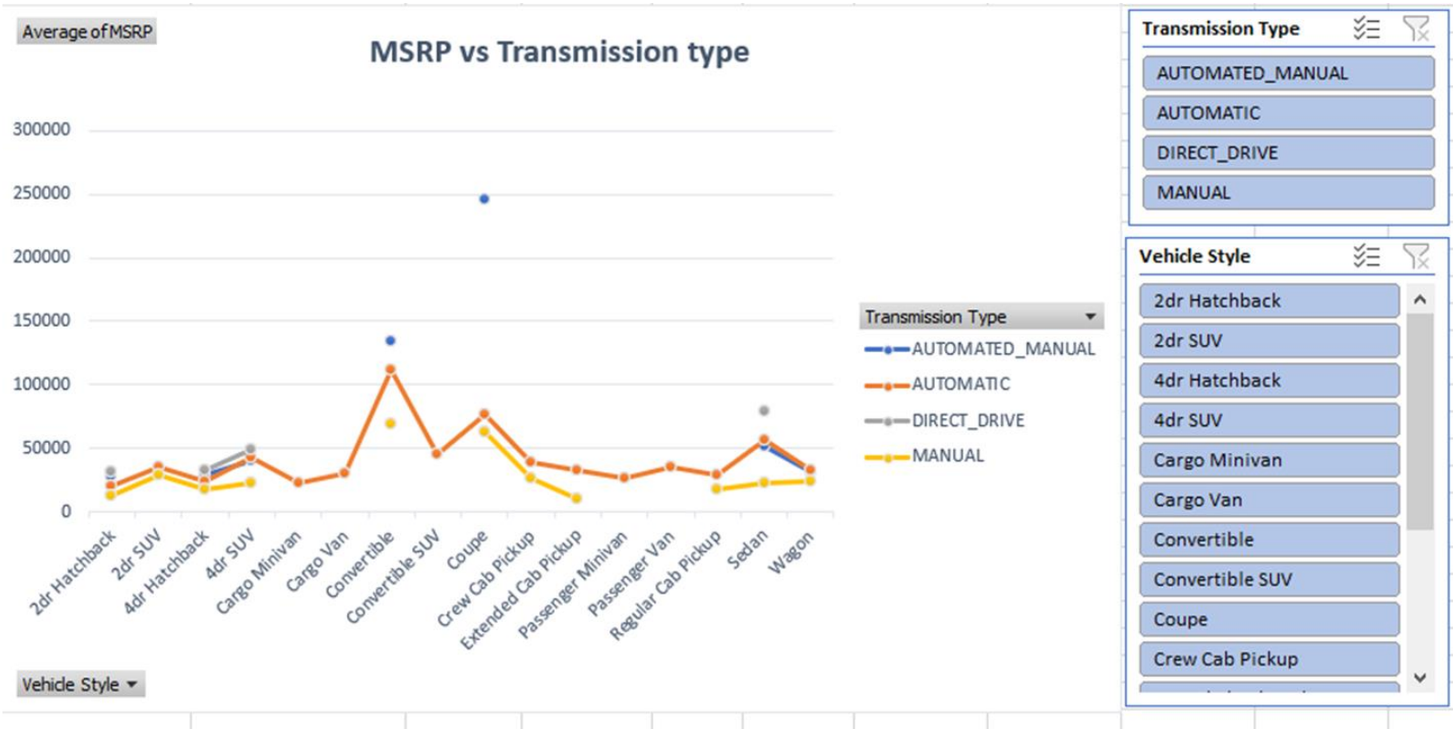
Task 1: How does the distribution of car prices vary by brand and body style?



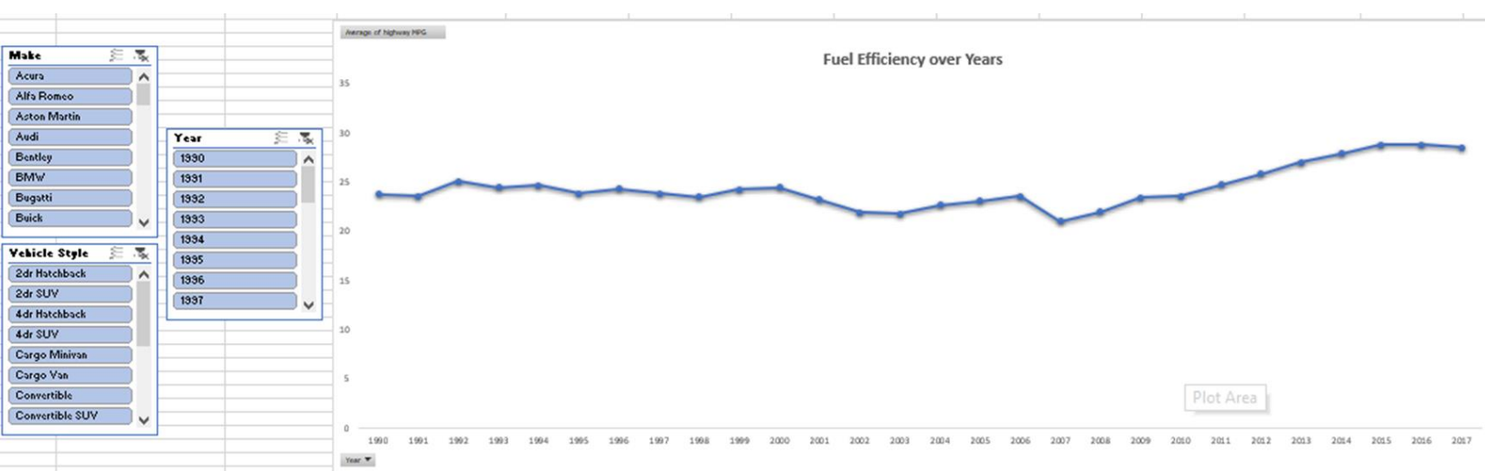
Task 2 :Which car brands have the highest and lowest average MSRPs, and how does this vary by body style?



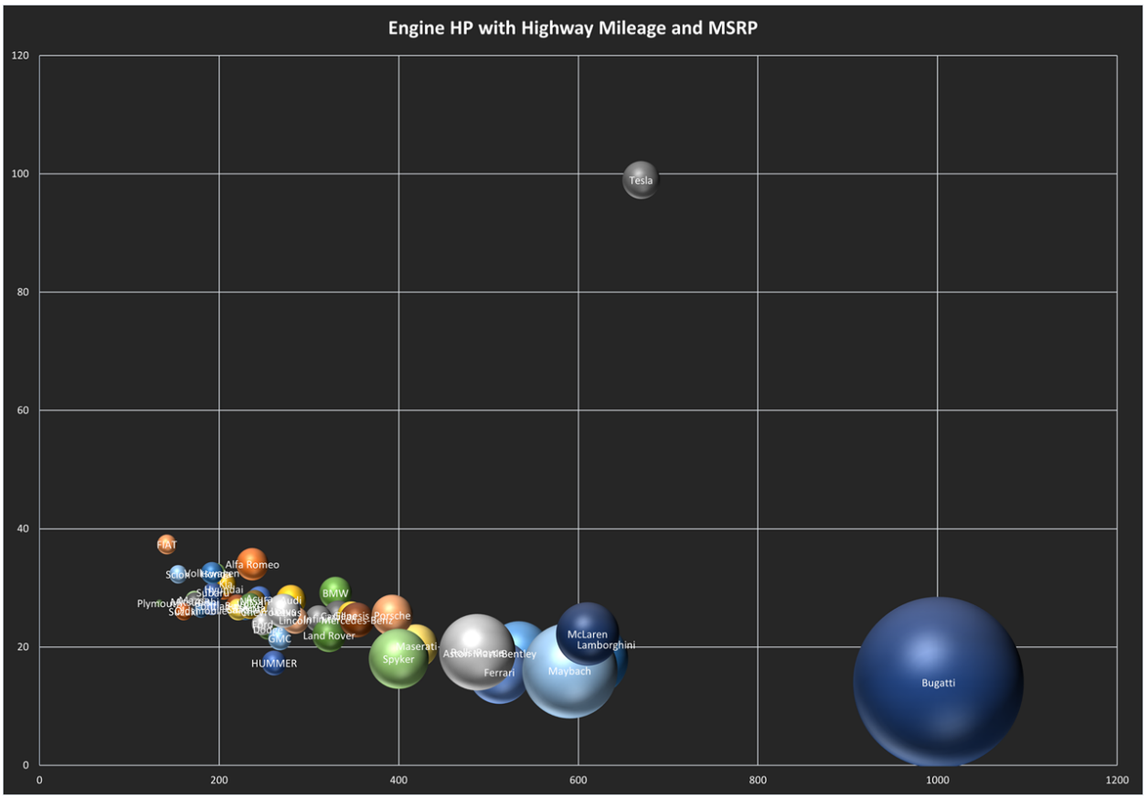
Task 3 : How do the different feature such as transmission type affect the MSRP, and how does this vary by body style?



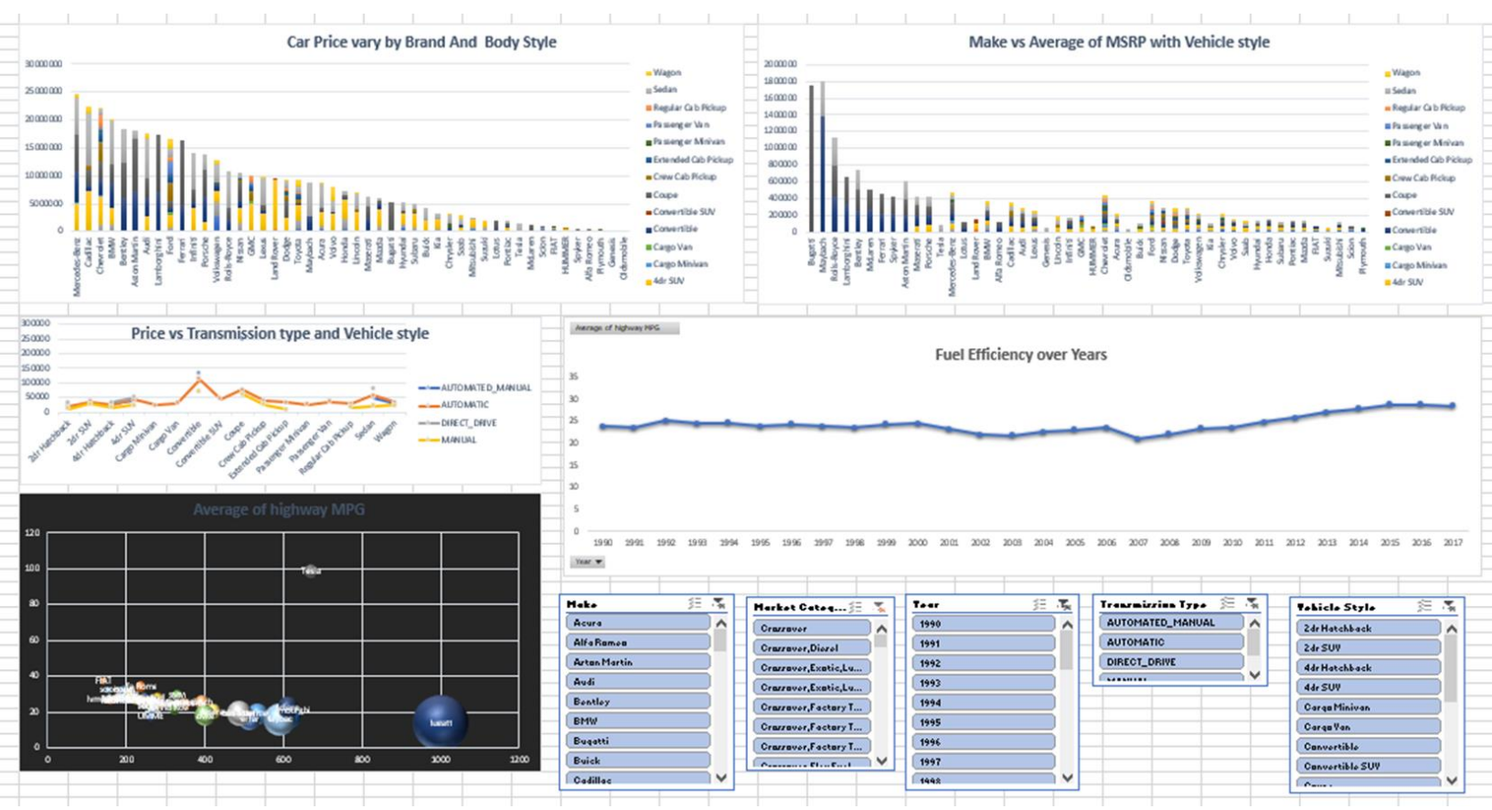
Task 4: How does the fuel efficiency of cars vary across different body styles and model years?



Task 5: How does the car's horsepower, MPG, and price vary across different Brands?



Final Interactive Dashboard



Insights

People in US like engine with higher horsepower. As the engine power increases, so does the price.

Engine Horsepower and Number of Cylinder a car has are the major contributors deciding the price followed by efficiency of the car.

Exotic cars hold up their value and aspiration over the years.

Electric cars tend to loose their value over the years.

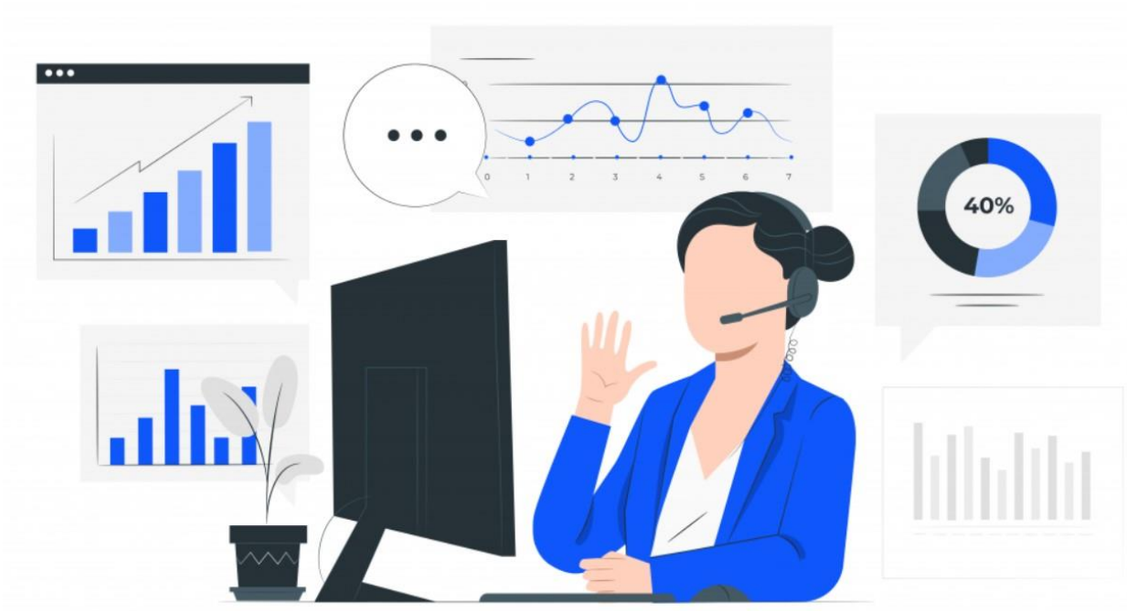
The efficiency of the car is inversely related to the Engine Horse Power and the Number of Cylinders. As the Power and No. of Cylinders increase, efficiency of the car decreases.

US market is different from Indian market, Indian market is more price conscious. In India, people usually look for efficient cars and do not mind compromising on Engine HP or how big the engine is. But in US, people like big powerful cars.

Result

This project involved extensive use of Excel. Pivot tables were extensively used to accomplish the tasks. Major challenge was to understand the data to fill the NULL/Missing values and the data irregularities rather than just dropping the NULLs as they would have dropped particular car brands all together and the pricing for those would be unknown by the dealers. Creating dashboard and Regression Analysis helped me understand the factors affecting the car prices better. Overall the project was challenging and intuitive at the same time.

ABC Call Volume Trend Analysis



Project Description:

ABC Call Volume Trend Analysis

Customer Experience (CX) Analytics for Inbound Calling Team

Overview:

This project focuses on analysing the performance of a company's inbound calling team over 23 days. We will look at details like how long customers wait before speaking to an agent, the length of the calls, and the outcome of the calls (whether they were abandoned, answered, or transferred). Our goal is to find useful insights to help the team handle calls more efficiently and plan staffing better. By improving these areas, we aim to enhance the overall customer experience and satisfaction. This analysis will involve calculating average call durations, understanding call volume trends, and making staffing recommendations.

Assumptions:

- An agent works 6 days a week.

- Each agent takes 4 unplanned leaves per month.

- An agent works 9 hours a day, with 1.5 hours for breaks.

- Agents spend 60% of their working hours on calls.

- A month has 30 days.

Objective:

The goal is to analyse call trends, improve agent efficiency, and enhance the overall customer experience by ensuring timely response to customer calls.

Tech-Stack Used:

MS Excel for analysis of data and visual representation of dataset. Adobe PDF for Presenting.

Approach and Insights:

Understanding Dataset:

The dataset covers 23 days and includes: - Agent's name and ID - Queue time (how long a customer waited) - Call time and duration - Call status (answered, abandoned, transferred)

Dataset has 117989 rows and 13 columns.

Data Cleaning:

• Filled blank values in wrapped by with abandon Call as it’s corresponding row data seems like the call was abandoned and no executive attended it.

Analysis and Insights:

1. Average Call Duration: Determine the average duration of all incoming calls received by agents. This should be calculated for each time bucket.

Your Task: What is the average duration of calls for each time bucket?

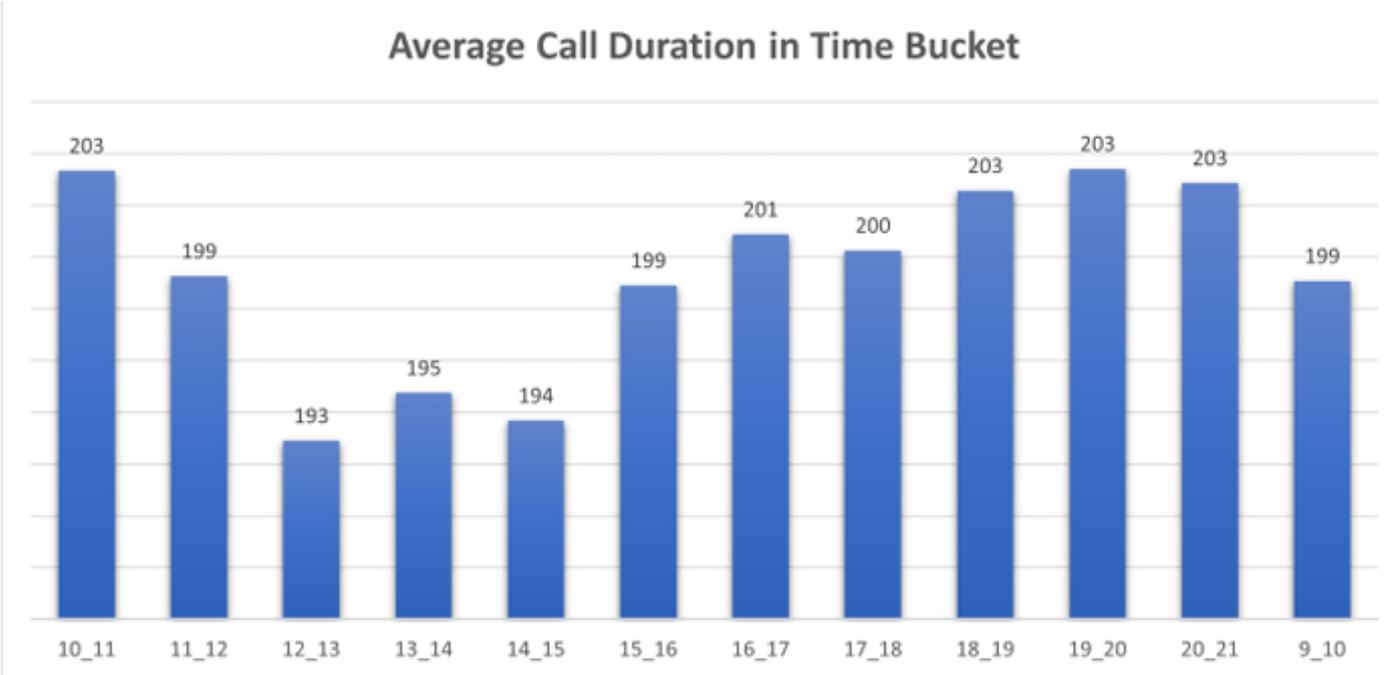
• Used a Pivot table to find the average duration of calls where the rows of pivot table are time

buckets and values are Average of Call duration in Seconds.

• Created a filter to filter out only answered calls.

• The total average call duration for incoming calls received by the agent is 198.62 seconds.

Call_Status	answered
Time Bucket	Average of Call_Seconds (s)
10_11	203
11_12	199
12_13	193
13_14	195
14_15	194
15_16	199
16_17	201
17_18	200
18_19	203
19_20	203
20_21	203
9_10	199
Grand Total	**199**



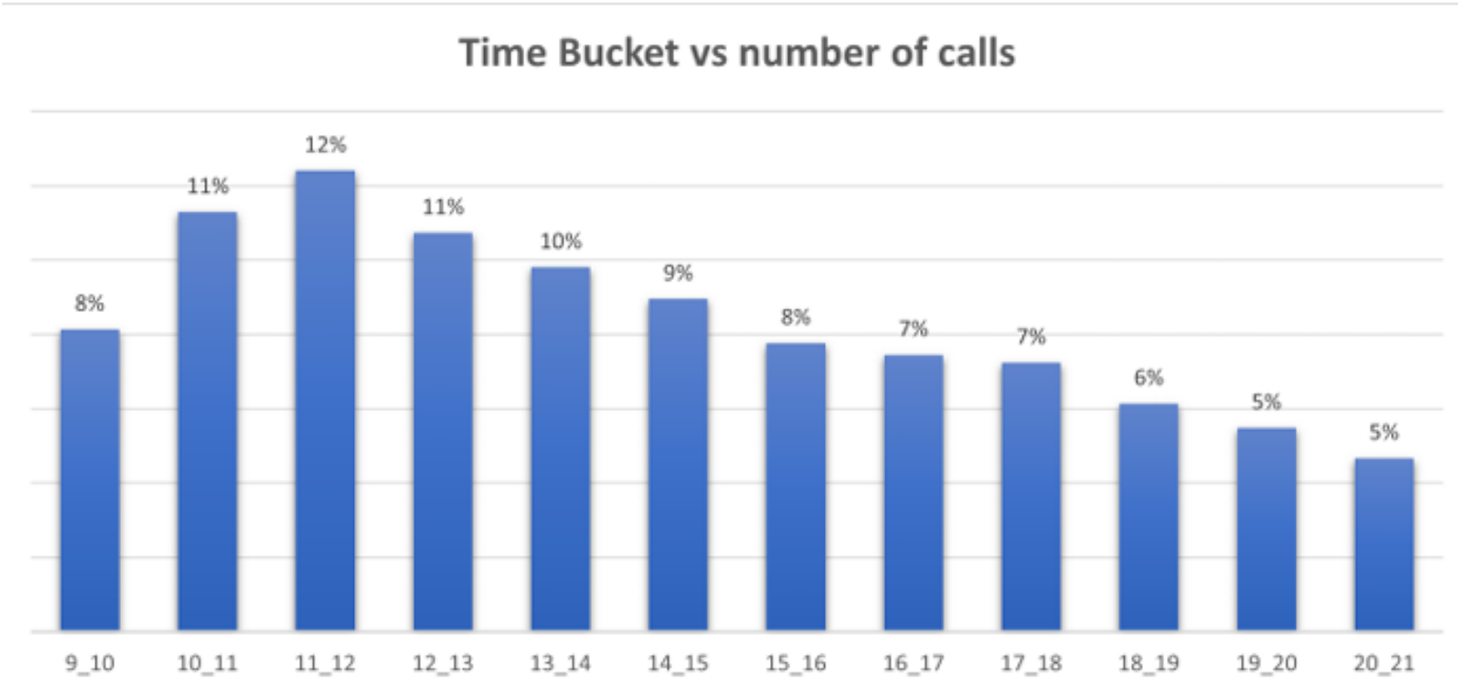
2. Call Volume Analysis: Visualize the total number of calls received. This should be represented as a graph or chart showing the number of calls against time. Time should be represented in buckets (e.g., 1-2, 2-3, etc.).

Your Task: Can you create a chart or graph that shows the number of calls received in each time bucket?

• Used a Pivot table to find the Call Volume the rows of pivot table are time buckets and values are Count of Customer Phone no.

Call_Status	All
Row Labels	**Count of Customer_Phone_No**
10_11	13313
11_12	14626
12_13	12652
13_14	11561
14_15	10561
15_16	9159
16_17	8788
17_18	8534
18_19	7238
19_20	6463
20_21	5505
9_10	9588
Grand Total	**117988**

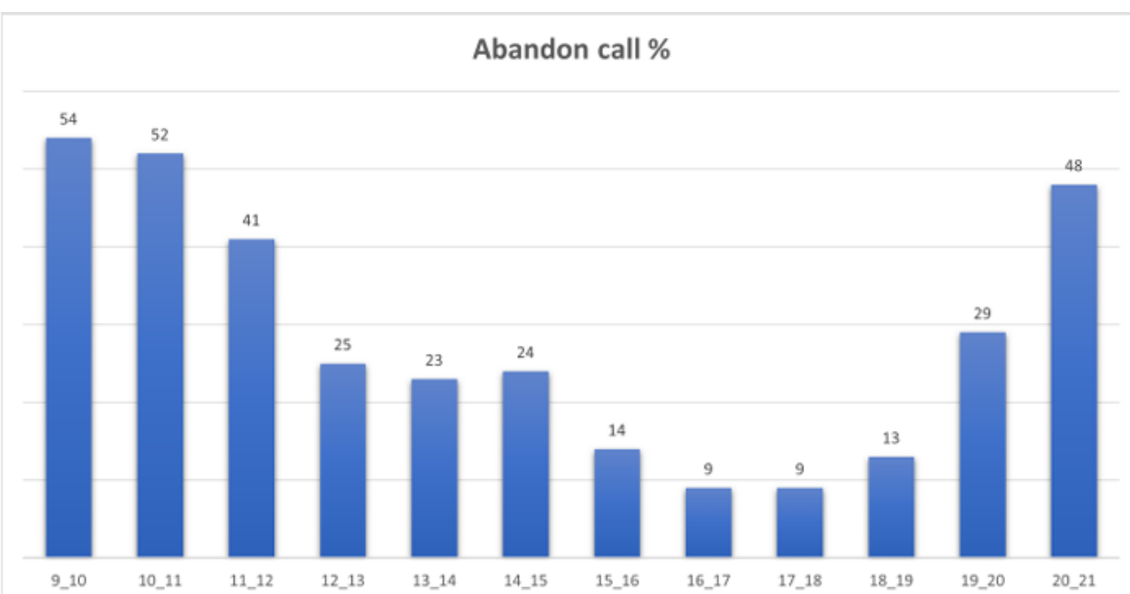
Chart Area



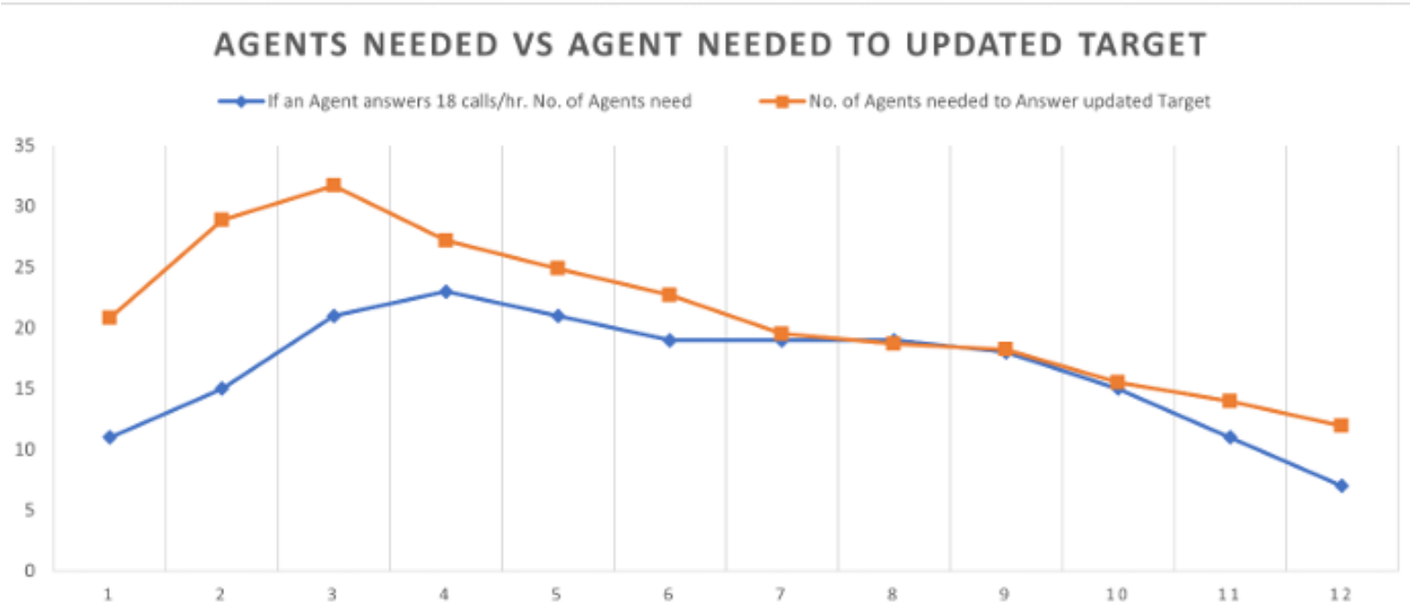
3. Manpower Planning: The current rate of abandoned calls is approximately 30%. Propose a plan for manpower allocation during each time bucket (from 9 am to 9 pm) to reduce the abandon rate to 10%. In other words, you need to calculate the minimum number of agents required in each time bucket to ensure that at least 90 out of 100 calls are answered.

Your Task: What is the minimum number of agents required in each time bucket to reduce the

abandon rate to 10%?



Highest Abandoned calls are 9-12 and 20-21.



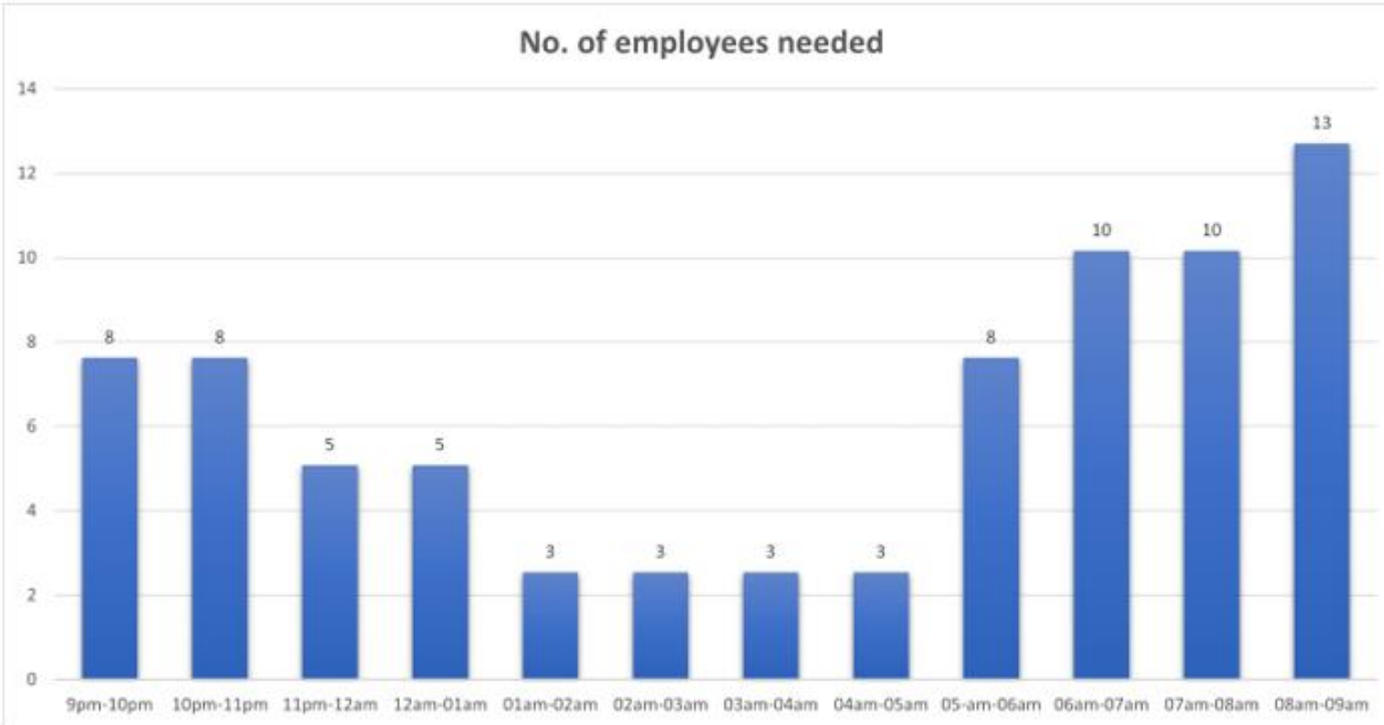
Agents needed are higher for updated target where there are very few agents when they answer 18 calls per hour.

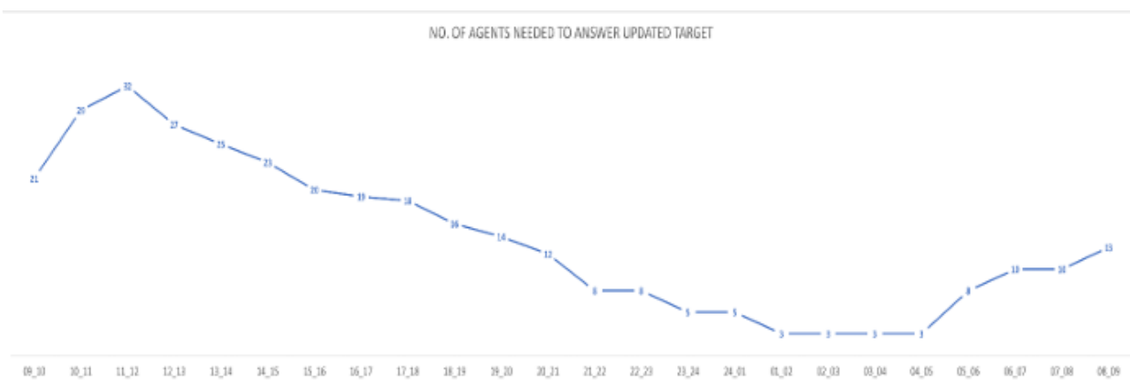
4. Night Shift Manpower Planning: Customers also call ABC Insurance Company at night but don't get an answer because there are no agents available. This creates a poor customer experience. Assume that for every 100 calls that customers make between 9 am and 9 pm, they also make 30 calls at night between 9 pm and 9 am. The distribution of these 30 calls is as follows:

Your Task: Propose a manpower plan for each time bucket throughout the day, keeping the maximum abandon rate at 10%.

Assumptions: An agent works for 6 days a week; On average, each agent takes 4 unplanned leaves per month; An agent's total working hours are 9 hours, out of which 1.5 hours are spent on lunch and snacks in the office. On average, an agent spends 60% of their total actual working hours (i.e., 60% of 7.5 hours) on calls with customers/users. The total number of days in a month is 30.

Distribution of 30 calls coming in night for every 100 calls coming in between 9am - 9pm (i.e. 12 hrs slot)											
9pm- 10pm	10pm - 11pm	11pm- 12am	12am- 1am	1am - 2am	2am - 3am	3am - 4am	4am - 5am	5am - 6am	6am - 7am	7am - 8am	8am - 9am
3	3	2	2	1	1	1	1	3	4	4	5





Proposal of Man Power Night Shift:

We need more agents during 5am – 9am and 9pm to 10pm.

The line graph above also shows the same but its railway time where more agents are required from 20 – 23 and 05 – 09 for updated target.

Note: the line chart contains 24hrs or data to meet updated target.

Insights:

• Average time duration of a call is 199 seconds.

• An agent on an average can handle 81 calls per day.

• Consistent average call durations across all time buckets suggest stable operations, uniform service quality, efficient call handling, consistent customer behaviour patterns.

• The maximum number of agents required is in the morning hours from 9 A.M to 1 P.M and the least number of agents required is at night hours from 11 P.M to 5 A.M.

• At present, no agents are available during the night shifts. Therefore, the company should either employ more agents for night shifts or consider changing the shifts of some agents.

Result

This project has aided me in understanding the importance of data analytics in customer experience analysis by providing valuable insights necessary for data-driven decision-making. Through this project, I gained insights into several aspects, including call abandonment rates, call duration distribution, call volume, agent allocation, and strategies for devising a workforce plan aimed at reducing abandoned calls.